The Report Committee for Bolin Lan

Certifies that this is the approved version of the following professional report:

From Inter-County Commutes to Megaregional Development in Texas

APPROVED BY

SUPERVISING COMMITTEE:

Ming Zhang, Supervisor

Alex Karner, Co-Supervisor

From Inter-County Commutes to Megaregional Development in Texas

by

Bolin Lan

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Community and Regional Planning

The University of Texas at Austin

August 2020

Acknowledgement

I would like to thank my supervisors Professor Ming Zhang and Alex Karner for their support and guidance. It is no doubt that I am unable to complete the project without their profound knowledge and insights on transportation planning. Especially for Professor Ming Zhang, his kindness helps me get it through during this difficult time.

Meanwhile, I also really appreciate my girlfriend, Ziyue Su, for her endless support and inspiring advices on my project. It has been an amazing wonderful two-year in Austin. Even though currently the COVID-19 pandemic is hitting hard around the world, I appreciate all the effort from faculties, colleagues, and friends to fight for our future.

Abstract

From Inter-County Commutes to Megaregional Development in Texas

Bolin Lan, M.S.C.R.P

The University of Texas at Austin, 2020

Supervisor: Ming Zhang

Co-Supervisor: Alex Karner

The study adopts network approach and measures the interconnection of Texas counties in an inter-county commuting network. Over the past decades, the Texas's commuting landscape has been drastically shaped and changed by the economic ties amongst different areas. Commute plays a pivotal role in structing the economic geography. More specifically, this study employs a community detection algorithm from the field of network science, namely the Girvan-Newman algorithm, to examine how 254 counties in Texas are connected through over 9,000 commuting flows at county level during the 2011-2015. The analysis utilizes accessible ACS 5-year (2011-2015) County-to-County Commuting Flows data set as an example to reveal the county or metropolitan clusters and indicated regional agglomeration in Texas based on the interconnection strength between nodes.

It was found that Texas has a dominant triangular commuting region, anchored by four large well-known metropolitan areas: Dallas-Fort Worth, Houston, San Antonio, Austin. Moreover,

the commuting clusters at county level characterize the economic interaction among major metropolitan areas in Texas. Reflecting on the results from the community detection partitioning algorithm, the study emphasizes on the importance and necessity of detecting county clusters for megaregional planning strategies.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Social, technological, and economic developments, which are largely influencing residents' activities and spatial patterns, are resulting in regions' considerable transformation across the world. It is no doubt that continued growth of urban areas is leading to economic interdependence between cities and their surrounding settlements (OECD, 2018). Apart from that, the increasing mobility and linkages between cities or metro areas through new technology advancement contribute to the new geographic scale of megaregions (OECD, 2018). According to Regional Plan Association (2006), the increasing connectivity between cities and regions can bring both new opportunities and challenges facing policy makers in cities and their neighborhoods, which are not supposed to be solved by action taken simply at city or metropolitan level.

Generally speaking, the term megaregion can be defined as a chain of interconnected cities or regions, which are geographically distinct but connected through trade, commute patterns, and transportation infrastructure. The concept of megaregions has received considerable critical attention within the field of urban planning. Back to 1957, Gottman firstly came up with the definition of metropolis or metropolitan, based on the concept of urban agglomeration. Urban agglomeration is generally recognized as one or more cities that are contiguously surrounded by and strongly associated with other big cities (Wu et al., 2019). These core cities are defined according to a specific population and they present the obvious functional dominance. Afterward, these buzzwords had been evolving under different social context. As a trans-metropolitan urban structure, the megapolitan, megalopolis or megaregion, has been once again back into focus by the success of European trans-national urban regions (Nambiar, 2006). Lang and Nelson (2007) explored the geography of "megapolitan areas", which are super regions combining at least two, and often several, metropolitan areas. Ross (2009) emphasized that “Increasingly, the most appropriate unit of social organization and economic coordination is not the city, not even the metropolitan area: it is the city-region or the region-wide network of cities”. Being similar with those highly integrated urban or city clusters, such metropolitan integrations can also bring numerous opportunities or challenges facing the planning field in the American context. The efforts of identification for megaregions in the United States continue to grow. Over the past decades, most of the national population growth with high rate concentrated in existing metropolitan areas in the United States, and it is no doubt that they are becoming more and more integrated in terms of economy, industry, or policy-making. One of the outcomes of those interconnected urbanized or metropolitan clusters could be the encouragement of daily commutes and flows (Weber, 2003).

However, such notion of megaregion, as a new scale of geography, has been still experiencing significant discussions among policy makers. Some of observers deem the emergence of megaregion as a phenomenon for granted, while others consistently emphasize on the importance and necessity of “megaregional” planning actions. Additionally, the question that what does consist of the real proof that megaregions are “coherent, powerful units of economic activity” is expected to be addressed. In Texas, it has been predicted that nearly 70% of the state’s population in 2050 will be concentrated within a triangular megaregion, coined as Texas Triangle. The form of the Texas Triangle is the result from dynamic transformation of the Texas’ four anchor metropolitan areas into an emergent megaregion: Houston, Dallas-Fort Worth, San Antonio, and Austin. A megaregion, officially proposed by America 2050 (Regional Planning Association), is a “large, connected network of metropolitan areas joined together by environmental, cultural, infrastructural and functional characteristics.” Unlike the other 10 megaregions in U.S., it is interesting for Texans to witness a triangular megaregion that is neither under linear nor contiguous development. Considering the unique characteristic, several questions could be posed that what makes this "Triangle" a functional mega-city? What might constitute real evidence that megaregions (in this case, Texas Triangle) are coherent, powerful units of economic activity? One appropriate answer lies in the high and increasing degree of integration found among their metropolitan areas' economies and societies, which can be evidenced by the economic, informational, and human flows among these four magnificent urban centers (Neuman & Bright, 2008).

Because the dominance of such a triangular megaregion in Texas is supposed to be fairly emphasized, this report comes into being by exploring a new interdisciplinary empirical approach employing the Decennial Census and ACS inter-county commuting flows data set as a proxy for patterns of economic interconnection to examine the county or even metropolitan clusters in the state of Texas. Detecting the clusters of counties over commuting network can help us discover the groups of counties that are highly connected, laying the solid foundation of the formation and development of mega-agglomerations(Yue et al., 2019). The exploration of these data sets allows to not only depict the interconnected nature of labor markets, but also demonstrate the constant integration of core metropolitan areas in Texas. Therefore, one of main objectives in the study is to provide compelling evidence-based assessments of mega-agglomeration in the state of Texas, by employing the interdisciplinary application of community detection algorithms from the field of network science. The prospective results should inform the decisions of policymakers and urban planners representing numerous initiatives from economic development to transportation planning in Texas, and prepare them for both challenges and opportunities in megaregional planning as well.

Other than simply handling with the descriptive statistics of the inter-county commuting flow data set and visualizing the origin-destination (O-D) flows, this study takes advantage of one network-partitioning algorithm, named as Girvan-Newman (GN). By recognizing every county of Texas as a node (or vertex) and any single flow as an edge connecting the nodes, the data set of inter-county commuting flows of Texas displays a complex "network". A network is supposed to have "community structure" if the nodes of the network can be easily classified into various groups, and each group contains a number of nodes that is densely connected internally. Even though Fisher and Ness (1971) claimed that it is not possible to find out the optimal clustering procedure, several computation algorithms for "community detection" have been developed and applied in network analysis with different levels of success. In this study, GN algorithm shows its utility and reliability in community structure identification for Texas with 254 nodes (counties) and over 6000 edges (one-way flows). Such an algorithmic method provides a regionalization scheme which totally depends on spatial laws, instead of contestable human interpretation (Nelson, 2016).

The remainder of the report is organized as follows. *Chapter 2* discusses the relevant literature on the application of the science of social network analysis in urban and regional planning, especially how community detection algorithms are applied to the planning field. In addition, in *Chapter 2*, the concept of urban agglomeration will be reviewed through previous studies. *Chapter 3* comes up with methodological framework for the study and how the comprehensive analytical process will be carried out. *Chapter 4* consists of several crucial findings with regard to the statistics, visualization, and algorithmic outcomes of the data set. *Chapter 5* summarizes the limitations and any suggestion for the current and future research, respectively. Conclusions are given in the *Chapter 6*.

Chapter 2: Literature Review

This chapter of literature review is divided into three sections. The initial part explores the emergence of megaregion and their importance as emphasized by diffe planning scholars. The second part discusses how increasing number of long-distance commutes or super commutes contributes to megaregional planning. The last part involves several empirical methods developed by researchers and scholars to delineate megaregional geography.

2.1 The Rise of Megaregions

In 1961, the French geographer Jean Gottmann firstly described a long-extended metropolis in the entire Northeastern US resulting from the growth of urban agglomerations between Boston and Washington. He termed it as "Megalopolis", within which the formation of a unique cluster of metropolitan areas extended beyond traditional and political borders (Gottmann, 1961). Inspired by Gottman's idea, Claiborne Pell (1966) began promoting transportation planning focused on this new scale of geography. He suggested that Megalopolis was becoming "entangled in ribbons of concrete," which should have more efficient travel such as intercity passenger rail service in the Northeast corridor. Additionally, The Region's Growth (Regional Plan Association, 1967) encouraged policies to protect the Northeast's historic "nucleated" structure in order to prevent the spread of the city.

In a more sophisticated from, the determining attribute for megaregions rests on their external and internal functional linkages (OECD, 2018). Harrison and Hoyler (2015) proposed that the term of megaregion is used differently relying on the actors and their contexts and what the goals are followed. In the U.S context, Scott (2002) uses the term "city-regions", following the

Gottmann’s concept, as “spatially overlapping or convergent urban areas with surrounding hinterland for the United States”. Afterwards, within the America 2050 project, Regional Planning Association (2006) attempts to consider geographic and economic connectedness for the term of megaregion, and they proposed RPA’s methodology for defining U.S. megaregions along with its strengths and weaknesses. Depending on that notion, Florida et al. (2008) recognized megaregions as an “integrated set of cities and their surrounding suburban hinterland across which labour and capital can be reallocated at very low cost” (Florida, Gulden and C, 2008, p.459). Ross (2006) identify megaregions in a similar way, which are “networks of metropolitan centers and their surrounding areas”, “spatially and functionally linked through environmental, economic, and infrastructure interactions” (Ross, 2009, p.1). Meanwhile, the case for megaregions was strengthened not only by scholars, but also student and faculty research projects since 2004 including: “Plan for America”, a graduate urban planning studio at the University of Pennsylvania School of Design; University of Michigan, University of Texas at Austin, Georgia Institute of Technology, Portland State University, Arizona State University, and etc. (Todorovich, 2009). Each of these universities has put their efforts into defining a specific megaregion with its geography, and highlighting both challenges and opportunities for megaregional coordination on public policy and planning issues (Todorovich, 2009).

The different definitions of megaregions or city networks have in common that they are defined as an integrated system of cities with their surrounding suburban hinterland. Recognizing or identifying potential megaregions can contribute to the cooperation across administrative boundaries with the purpose of supporting economic competitiveness and addressing common challenges at the appropriate geographic scale (OECD, 2018).

2.2 The Growing Trends of Long-commutes or Super-commutes

As the telecommunication advancement is unstoppable, the place of work is no longer fixed in one location, but rather where the worker is situated. Consequently, people have moved outwards to suburban areas, and city labor sheds (where work live) have expanded over the past decades. This results in greater economic integration between cities or metro regions situated hundreds of miles apart. People are becoming more willing to commute long distance by air, rail, car, bus, or any combination of modes. Growing number of studies have discussed the impacts of long-distance commutes, or super commutes. The most popular study in the U.S. context belongs to Moss and Qing (2012), which examined the emergence of the "super-commuter" over the past decade. The analyzed the growth trend of super-commuting in major metropolitan regions, combining geographic and demographic characteristics. As a result, they affirmed that the expansion of city labor sheds across the states exemplified how the economic geography of American cities has evolved in this information age. Social and economic activities become increasingly inter-regional (Moss and Qing, 2012). Moreover, the shift towards "mega-regional" planning and much closer economic cooperation between cities, would undoubtedly apply to regions such as Arizona "Sun Corridor", the "Texas Triangle", and in California, all of which are well-established super-commute corridors (Moss and Qing, 2012).

The trends towards urban integration and "super-commuting" are not necessarily limited to the United States. To gain global competitiveness, nations across the world would love to address on the increasing trend of super-commuting. For instance, in the European context, Sandow and Westin (2010) identified the characteristics of individuals with different durations of long-distance commute and the factors incentivizing those commuters in Sweden. Furthermore, they analyzed the effects in terms of economic outcomes of long-distance commuters. They came up with the conclusions that long-distance commuting is a form of mobility affecting the lives of relatively high proportion of the Swedish workforce and their households. Identified the general relationships between long-distance commute and socio-economic factors. And finally helped transportation planning and policies reflect on a sustainable transport system, in terms of economic, environmental as well as social consequences of commuting. This study exclusively explored the specific relationships between duration(years) of long-distance commuting and several variables, such as education, age, income etc., which enlightened transportation agencies and make them reconsider the improvement of sustainable transportation system. But it did not consider the environmental effects of long-distance commuting even though it highlighted the importance of sustainable transportation system in the end of the article.

Anderson et al. (2018) contributed a similar examination regarding the growth and significance of long-distance commuting, characteristics of the commuters and the pathways to becoming a long-distance commuter in Sweden. They focused more on rural to urban long commuting. Conclusion was that there were two types of pathways to rural long-distance commuting: living and being employed in a rural municipality and becoming employed by an employer located in an urban region, and also, living and being employed in an urban region and keeping employment in an urban region while moving to a rural municipality. Working population living in city municipalities had been increasing over the period 1990 – 2009 in Sweden, while those in rural areas has been decreasing. Rural commuters constituted the largest fraction of all long-distance commuters, but the growth of it from 1990 and 2009 has been lower than the average. This paper could be one of a few studies exclusively focusing on rural to urban long-distance commuting against the backdrops of rapid urbanization. The paper provides empirical support for the argument that there has been a quite substantial increase in employer–employee ties between rural workers and urban employers. There are some overlapped and similar analysis with the previously reviewed paper, hence more innovative methodologies should be considered.

2.3 Empirical Approaches for Delineating Mega-Agglomeration or Megaregions

The term agglomeration is related to clusters of population or business activity. Urban agglomeration is the result of economic agglomeration within a region. As the skeleton of regional structure, the flows of people, goods, finance, and information are the fundamental connections between cities (Démurger 2001). Hence, commuting flow can be recognized as an important indicator to evaluate the development conditions for a region (Li et al. 2016). Within the literature there has been much research surrounding concepts of "commuting flow" and a large body of them discussed the importance of commutes in structing the economic geography. Flow data collected by national statistical institutes extends in most cases to commuting data. This type of data is usually analyzed for the identification of people's daily activity patterns and economic connectivity based on travel to work information (OECD, 2018). In order to fully explore this type of data that can be considered as being structed by nodes and edges, scholars put efforts in developing "community detection" or "community partitioning" algorithms from the field of network science.

Nelson and Rae (2016) provided a new perspective on how to redefine functional economic geography of the U.S. The goal was to partition the United States into functional megaregions. In the meantime, it shed light on the traditional way of dividing space into areal units, which has been a vexed problem in the field of geography over decades. The study used two approaches. In the first approach, a visual representation interpreted data spatially, saving time to identify anomalous connections. However, it was only a visual inspection alone, without the demonstration of statistical accuracy. In the second approach, algorithmic analyses were developed for a partitioning. Combo was used as the main tool for the algorithmic calculation and "modularity" score was counted as the main criterion to decide the number of partitions. Finally, the contiguous United States was able to be divided into geographically – contiguous regions. These regions are "interpretively recognizable as megaregions with major cities at their centers". The study innovatively combined "visual heuristic approach" and "algorithmic approach" to successfully redefine the functional economic geography of U.S. and identify latent "megaregions". Several runs of data were set to compare the different outcomes until the optimal one emerges and successfully enlightened transportation agencies and made them reconsider the regional transportation planning much carefully. Despite the success of the study in certain aspects, it still suffers from its "filtering" works in the algorithmic approach, which exclude tract pairs that are indeed spatially far away from each other. Strong "connections" may exist in excluded tract pairs, which is likely to contribute to a different visual delineation and algorithmic partition. While commuting data can actually deliver some interesting insights on the connectivity within and between cities and regions, simply depending on the results of this type of data should be carefully interpreted.

Inspired on Nelson and Rae's approach, Wu et al. (2019) took the nationwide car-hailing data set to delineating regional economic geography in China at different scales by examining the interconnection strength between nodes. They employed the community detection algorithms, namely Louvain algorithm, with adjustable resolutions and Combo with high-precision output, respectively (Wu et al., 2019). Even though both of the two algorithms ignore the spatial proximity, the study found that most of the prefecture-level cities in China have a dominant region with two to three smaller commuting sub-clusters, while regional central cities were extending their commuting hinterlands across jurisdictional boundaries. They affirmed the feasibility and reliability of community detection partitioning algorithms in the application of regional science, which should be more widely used in regional delimitation supported by big data. However, in order to furtherly improve the research, there is a need to explore new algorithms that can set up more parameters, such as accuracy and spatial distance.

Yue et al. (2019) attempted to evaluate whether the public intercity ground transportation infrastructures and services (here for passenger trains and long-distance buses) back up the integration and development of urban agglomerations in mainland of China. More specifically, the study concentrated in studying the strengths of transportation connections among the cities within urban agglomerations and detecting the clusters of cities over transportation network. The study employed a transportation cluster (TCD) approach combining K-shortest paths, hierarchical clustering, and geo-modularity. The datasets of passenger trains in mainland China and long-distance passenger buses were utilized in the study. As a result, each of the 13 urban agglomerations assigned by the Chinese government has at least one transportation cluster, and the core cities of each urban agglomeration are strongly connected. By overlaying the transportation clusters with the urban agglomerations delineated by the Chinese government, well connected and integrated agglomerations were identified, while some agglomerations were split into multiple transportation clusters and/or isolated cities. The results, meanwhile, revealed that in western China, where the economic development and urbanization are slower compared to the east, the ground transportation infrastructure and services are considerably weaker.

It is notable that a brand-new approach was proposed and successfully applied to the analysis, which inspired other scholars when conducting related transportation network analysis. The core cities and fringe cities of the clusters were explicitly identified, pointing out the main direction, objective and aim for transportation planners and decision-makers. However, the study has neglected, as authors' self-critiques, the actual flows between cities. The volume of the flow would reflect the strength of connection in a more realistic and accurate way, even though the public ground transportation infrastructures and services for urban agglomerations were evaluated. In addition, the study did not explicitly point out for those cities locating in transportation clusters but not in current regions agglomeration.

A more comprehensive overview for empirical approaches of delineating megaregions or mega-agglomeration comes from OECD's (Organisation for Economic Co-operation and Development) report in 2018 regarding the definition of megaregions. The author summarized that the approach used to delineate megaregions can derive from "morphological, functional or network " analysis according to previous scholars' works (Marull, Font and Boix, 2015; Ross, 2009): The morphological approach relies on continuous urban settlement areas, which may "reach specific thresholds of density, dimension, or degree of urbanization"; while in the functional or network approach, a megaregion will be identified as "an area of interactions between actors that can go in multiple directions and on several interconnected multiple layer" (OECD, 2018). Hence, identifying complex structures by functional or network approach asks for information on flows between different parts of the megaregion. "Commuting flows" or "commodity flows" can help with providing such information.

The author developed a mean-shift clustering algorithm to measure the megaregions on an international scale. Based on network theory, nodes refer to sample cities refer and edges to physical infrastructure that connect these cities (e.g. roads). "Megaregions are therefore characterized by the location of cities in space and are related to each other based on infrastructure links and distances" (OECD, 2018). The results presented that across the world, 25 megaregions with at least 10 million inhabitants living in the functional urban areas are identified by applying the mean-shift clustering algorithm.

It is obvious from the previous studies that the megaregional delineation will play as a crucial role to regional studies and more systematically empirical methods are supposed to be developed. New economic geographies require adjustments in the policy framework. In order to honor and maximize the economic and social benefits of megaregions, planning strategies at megaregional scale would be more than necessary at this stage.

Chapter 3: Study Methodology

3.1 Data Sources: County-to-County Commuting Flows Datasets

The Census Bureau provided a series of tables that include county and minor civil division (MCD)-based worker flow counts for the U.S. and Puerto Rico. The study mainly uses the accessible table "Residence County to Workplace County Flows for the United States and Puerto Rico Sorted by Residence Geography: 2011-2015". In the survey, full addresses of a worker's residence and workplace are collected. The ACS, in the meantime, emphasizes that collected place-of-work data shows some workers who made atypical daily work trips (e.g., workers who lived in New York and worked in California), due to people who worked during the reference week at a location that was different from their usual place of work, such as people away from home on business (ACS Design and Methodology Report, 2014). Prior to the ACS, the Decennial Census was developed for producing county and MCD-based level commuting flow tables and the 1990 Census is the earliest one. In the first part of the study presenting the general statistics and visualization of flow mapping, the 1990 Census table is utilized as well to present the changing economic geography visually and statistically in Texas compared to the table of ACS 2011-2015.

Counties are used as basic unit of the analysis for this study. The commuting flow files provide the origination and destination counties associated with the county FIPS. The total number of commuters in each flow is also recorded. A representation of the data of origin-destination for 1990 and 2011-2015 is given as below in *Table. 1*. By picking out the counties within Texas, the final data set fitted in this study for GN algorithm contains 6,370 county-to-county commuting flows in Texas, associated with a total of 2,583,103 commuters in 254 counties.

Residence				Place of Work				Commuting Flow	
State FIPS Code	County FIPS Code	State Name	County Name	State FIPS Code	County FIPS Code	State Name	County Name	Workers in Commuting Flow	Margin of Error
48	001	Texas	Anderson	048	001	Texas	Anderson	15,925	687
48	001	Texas	Anderson	048	005	Texas	Angelina	4	6
48	001	Texas	Anderson	048	029	Texas	Bexar	41	40
48	001	Texas	Anderson	048	037	Texas	Bowie	30	48
48	001	Texas	Anderson	048	041	Texas	Brazos	54	55
48	001	Texas	Anderson	048	067	Texas	Cass	11	12
48	001	Texas	Anderson	048	073	Texas	Cherokee	480	208
48	001	Texas	Anderson	048	085	Texas	Collin	4	7
48	001	Texas	Anderson	048	113	Texas	Dallas	102	70
48	001	Texas	Anderson	048	141	Texas	El Paso	7	10
48	001	Texas	Anderson	048	147	Texas	Fannin	10	16
48	001	Texas	Anderson	048	159	Texas	Franklin	14	23
48	001	Texas	Anderson	048	161	Texas	Freestone	193	80
48	001	Texas	Anderson	048	183	Texas	Gregg	110	59
48	001	Texas	Anderson	048	201	Texas	Harris	179	106
48	001	Texas	Anderson	048	203	Texas	Harrison	10	17
48	001	Texas	Anderson	048	213	Texas	Henderso	293	150

Table. 1. Commuting Flow Table Representation

3.2 Methods

Two approaches are employed in the study: one needing general visual interpretation and exhaustive comparison of descriptive statistics, while the other relying on algorithmic computation. In the former method, the visualization of flow mapping can identify the strongly interconnected county pairs and their patterns of spatial clustering. Meanwhile, the study conducts a comparison analysis between flow mappings of 1990 and 2011-2015 as its first approach to examine Texas regional delineation. Additionally, the analysis of super-commuting in Texas is conducted as well. This method helps with providing a general picture of the changing trend of commute patterns over the past decades. In order to present an effective visualization of how the commuting landscape has changed over the past two decades, the specific flow volume thresholds are set. Each desireline (one-way flow, from residence to workplace) finally represents a certain number of commuters by being assigned a specific color. Consequently, for both mappings of 1990 and 2011-2015, clear-cut presentations of commuting patterns and economic geography of Texas were yielded by keeping desirelines on the map that each of them has volume of over 100 commuters. Additionally, the results from the statistical comparison between two spans are similar.

Another approach has been employed for regional delineation is the algorithmic computation. Many systems of scientific interest form as networks, bunches of nodes or vertices that are joined together in pairs by links or edges. Typical examples are social networks, technological networks food webs and etc. (Girvan and Newman, 2002). Finding “community structure” in a given network serve as one of the best data analysis techniques used to recognize the structure of network datasets in large scale (Girvan and Newman, 2002). A “community” in a network is a set of nodes in which the density of connections is stronger internally with the community than it is externally with members of different communities (Nelson, 2016). In the first approach of visual presentation, the flow maps can be rendered as nodes or vertices that are joined by many lines, and those communities are presented as nodes or vertices with dense connection by lines. However, the ease of visual interpretation for communities does not meet the mathematical definition, which relies heavily on computational calculation of dense connections rather than spatial positioning of the nodes. Considering the structural similarities, it is reasonable to treat the

ACS commuting flow dataset of Texas (2011-2015) as a network consisting of nodes (Counties in Texas) and edges (commuting flows). Therefore, the network in the study consists of 254 nodes and 6370 edges. *Figure.1* illustrates the structure of commuting network without considering spatial proximity.

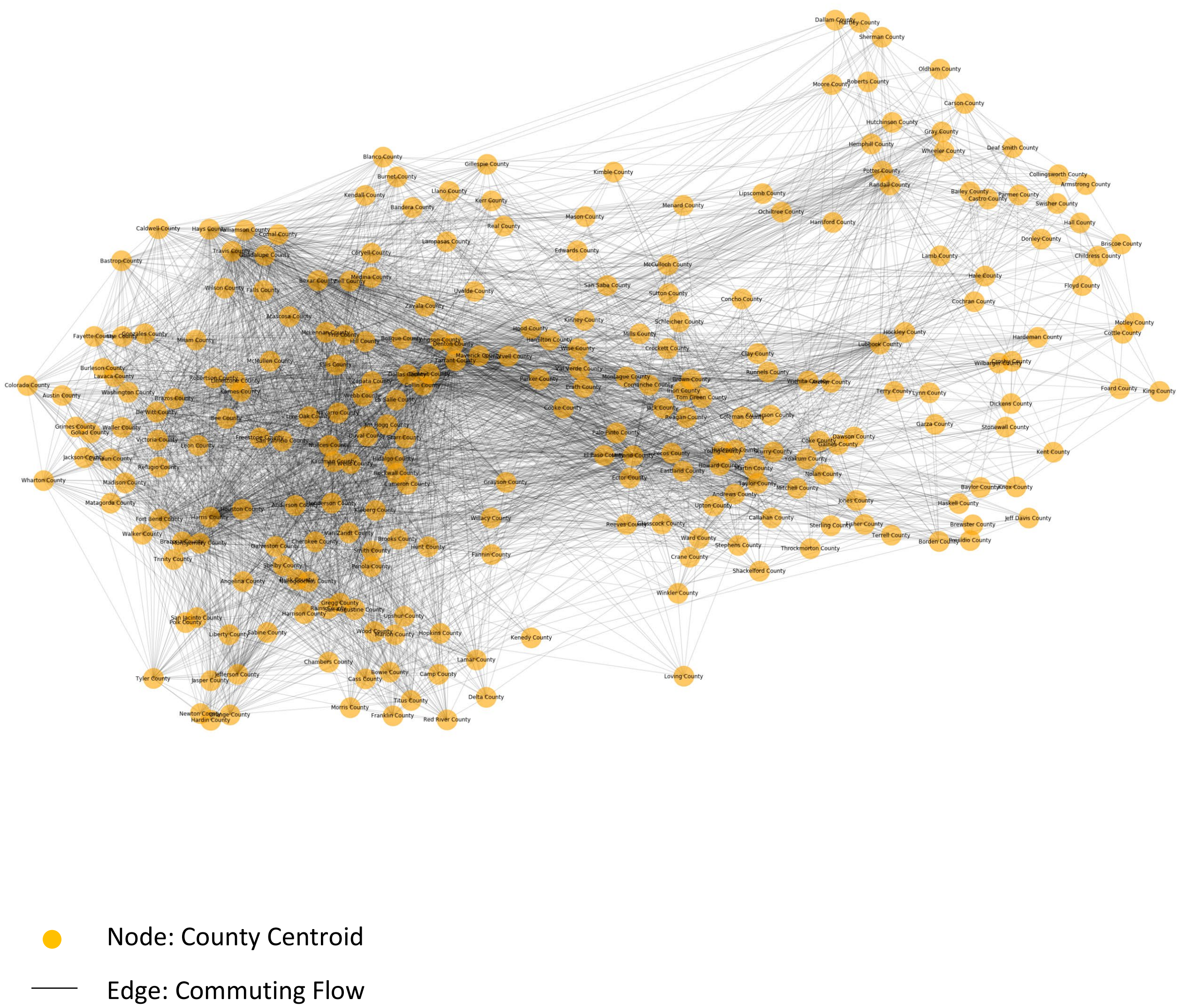

Figure.1. Inter-County Commuting Network of Texas without Defined Spatial Proximity

3.2.1 Divisive Hierarchical Clustering

In network analysis, detecting communities in a network functions as one of the most important tasks. One good example is an online social network consisting of millions of nodes and edges. Detecting communities in such network becomes a herculean task. Therefore, numerous well-developed algorithms for detecting communities over a certain network emerged in the past 20 years. One of them lies in the hierarchical clustering technique. Typically, this is an algorithm that can divide the objects into different groups or clusters. The end of this algorithmic computation is a bunch of clusters, where each of them is distinct and the objects within each cluster are at high level of similarity. Furthermore, based on the vertex connectivity or similarity, hierarchical clustering algorithm can be divided into two types: agglomerative and divisive. The agglomerative clustering is a "bottom-up" algorithm that each element is initially clustered and the two clusters will be combined into a new larger cluster, which the divisive clustering is the "top-down" one that assigns every element into a single cluster by initially setting up all elements into one same cluster. Newman and Girvan (2002) criticizes the agglomerative hierarchical clustering could be useful but far from perfect and demonstrates its failure in finding the community structure which is well-known from other studies. Therefore, a divisive hierarchical clustering method namely Girvan-Newman algorithm (GN), proposed by Girvan and Newman themselves, is employed in this study. This GN algorithm depends on repeatedly removes the edges of the maximum betweenness centrality in the network and by getting rid of the edges, the network will successfully break into smaller structures - communities. The edge betweenness centrality in a network is described as the number of shortest paths that go through an edge, which is also a type of attributes that belong to that edge.

Within a network, the shortest path means the path between any two nodes that covers the least amount of distance (or weights in a weighted network). *Figure.2* illustrates in a simple way that how many shortest paths that each edge is supposed to have in the given network. In this example of simple network consisting of 10 nodes, the degree of betweenness centrality for edge EF is 25 and this edge represents as a bridge-like connector between two parts of this network, (A,B,C,D,E) and (F,G,H,I,J). The removal of edge EF will result in a partition of this network into two well-connected clusters.

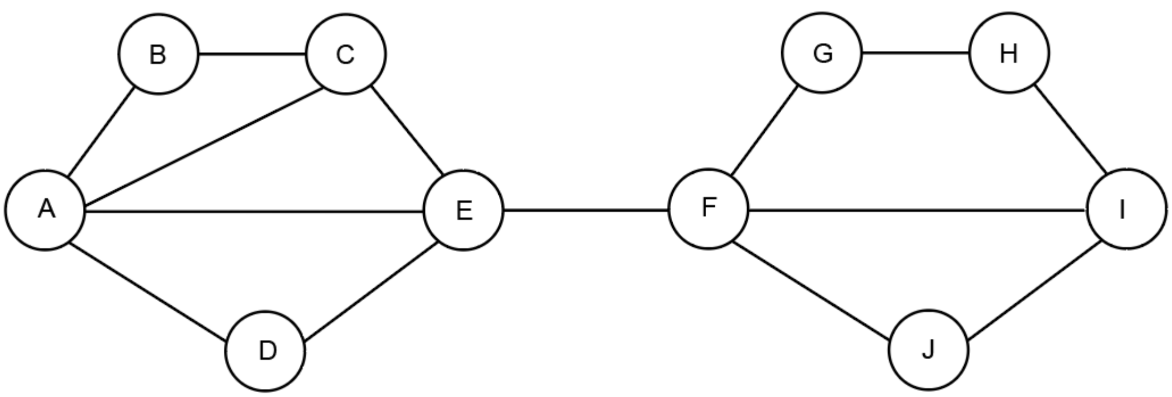

Figure.2. Edge Betweenness Illustration

More specifically, GN algorithm proceeds as following steps (Girvan and Newman, 2002):

1. Calculate the betweenness for all edges in the network.

2. Remove the edge with the highest betweenness.

3. Recalculate betweennesses for all edges affected by removal.

4. Repeat from step 2 until no edges remain.

3.2.2 County Cluster Detection

Because of the nature of GN algorithm, it is a divisive hierarchical clustering which needs a criterion to evaluate the division result of communities or clusters. Newman and Girvan (2004), afterward, had introduced a quantitative index, named as Modularity, which computes the difference between the number of edges within communities and the expected number in a random graph or a network. They mentioned in the study that "a good partition of a network should result in a significantly greater number of edges within communities than expected". The modularity is expressed mathematically as follows (Newman and Girvan, 2004):

$$Q = \sum_{i=1}^{k} (e_{ii} - a_i^2)$$

Where e_{ii} is the fraction of edges in the given network connecting nodes in the same community i, and a_i is the fraction of edges which are attached to the nodes in community i. Higher modularity corresponds with more edges within the module that should be expected by chance. The more detailed mathematical expression of modularity is shown as follows:

$$Q = \frac{1}{2m} \times \sum_{vw} \left[A_{VW} - \frac{k_v k_w}{2m} \right] \times \delta(C_v, C_w)$$

where m is the number of edges; A_{vw} is the element of the A adjacency matrix in row v and column w; k_v is the degree of node v, the number of connections attached to the v-th node; k_w is the degree of node w, the number of connections attached to the w-th node; $C_v$and $C_w$are the types of two nodes (v and w); delta (x, y) is 1, if x = y, otherwise it is 0:

In this study, however, the edges' weights are supposed to be taken into account. The linkage coefficient (C_{ij}), with reference to Kongari et al. (2010), is the only variable to be considered as the weight of each edge, denoted as W_{ij} . For the calculation of W_{ij}, the number of in-commuters and out-commuters for each pair of counties and the total number of commuters in each of the counties are the variables. The linkage coefficient between two counties is finally defined as the total number of commuters between a pair of counties, divide by the total number of commuters in the two counties. This definition of a linkage coefficient quantifies the strength of the bond between any two counties (Kongari et al., 2011). The formula is given as below:

$$W_{ij} = T_{ij} \,/\, (T_i + T_j - T_{ij})$$

Where W_{ij} is the linkage coefficient between two counties i and *j*; T_{ij} denotes the number of commuters between two counties i and *j*; T_i is the total number of commuters into and out of county *i*; T_j is the total number of commuters into and out of county *j*. Of note, the link between is stronger with larger value of W_{ij}. Hence, by considering the weight of each edge, the formula of the weighted modularity should be modified as below:

$$Q = \frac{1}{2m} \times \sum_{ij} \left[W_{ij} - \frac{k_i k_j}{2m} \right] \times \delta(C_i, C_j)$$

W_{ij} remains as the weight of the corresponding edge *i-j*; k_i is the total weight of edges attached to vertex *i*, and kw is the total weight of edges attached to the vertex *j*; *m* is the total edge weights in the given network.

For all the calculation by GN algorithm, there are two analytical tools that are possibly fitted in this study, Gephi and Python programming. Gephi is an open-source and power for tool designed for network exploration, analysis, and visualization. But the output of GN algorithm in Gephi depends on the specific input parameters. For instance, what is the weight of each edge and how many clusters or communities are supposed to be input. Suppose that the structure of the ACS commuting flow network is unknown, those parameters needed for GN calculation in Gephi are unavailable. Hence, Python programming in this study is considered as a more manipulated tool. In Python, there is a package, namely NetworkX, for the "creation, manipulation, and study of the structure, dynamics, and functions of complex networks" (NetworkX, 2019). In order to build a network in NetworkX, the O-D nodes are numbered by FIPS code of each county and each flow between any O-D pair is designated as an edge. The results calculated by GN algorithm includes the number of communities alongside the community number that each node belongs to. For each number of communities, the correspondent modularity will be calculated as well. Finally, ArcGIS is used to match and visualize all the data and results. One thing needs to be mentioned that even though the GN algorithm has shortcomings in the network analysis, such as the long running time for the network of large size, and when the amounts of nodes in a given network exceed a certain number, the algorithm might not be applicable.

Chapter 4: Results

In this chapter of the report, the results of visual comparison and algorithmic analyses are presented. In the first part, the flow mappings of 1990 and 2011-2015 help with identifying natural “communities” or regions that are strongly connected where commutes take place. Depending on these, it is not difficult to visually find out the interconnected county pairs and the strength between them. Texas, as a result, is dominated by four metropolitan areas mentioned as the beginning of the report in both 1990 and 2011-2015 period. A triangular commuting structure crops up as expected. The flow mapping of 2011-2015 period even shows a picture of more strongly interconnected metro areas compared to 1900’s. All of these enable us to understand more about the underlying economic geography of major metropolitan areas of Texas.

The second part of this chapter is comprised of some interesting findings from the algorithmic method, which assigns each county into potential communities based on their relational positions in the network dataset. The prospective results deliver a more persuasive message: the economic integration of large metropolitan areas of Texas is unstoppable.

4.1 Identification for Triangular Spatial Structure

Over the past two decades, the total number of two-way flows has been decreased from 4,744 in 1990 to 4,668 during 2011-2015. In 1990, the largest inter-county flow (two-way) comes to between Dallas County and Tarrant County, followed by the flow between Collin County and Dallas County. During 2011-2015, the largest one switches to flow between Fort Bend County and Harris County, followed by the flow between Collin County and Dallas county.

In order to visually obtain a general picture of commuting flow datasets and check out the trend of commuting patterns, *Figure.3* and *Figure.4* was created as below to intuitively show the change of commuting flows between specific OD counties during the targeted time span. For both flow mappings, each flow represents the round-way commutes and various colors denote different volumes of commuters between any county pair. It is easy to see what appear to be a number of separate functional economic zones. For example, the flow mapping of 1990 shows several large interconnected urban regions existing around four largest cities in Texas: Dallas, Houston, San Antonio. During 2011-2015, another large commuter region could be observed which takes in the metropolitan areas of Austin-Round Rock. Additionally, during 2011-2015, the connectivity among these separate economic regions has considerably increased and a more complex representation of a triangular spatial structure emerges. Also, the economic connection between the western and the eastern state has been strengthening reflected by the flow mappings.

The initial plotting of inter-county commutes is useful because it provides a simple visual depiction of economic linkages and helps us understand the spatial structure of commuting in Texas over the past decades. It is noticeable that these patterns do map onto underlying patterns of population density, but they also convey other valuable information with regards to the connectivity of these areas with each other. However, the assumptions based on visual representations are imprecise and somewhat subjective. It is difficult to only know more about the underlying network structure of the inter-county commuting data if only relying on the mapping of flows alone. Therefore, algorithmic approach is necessary for real-world application, where the statistical accuracy is required. The following sections provide the statistical and algorithmic results from the data sets.

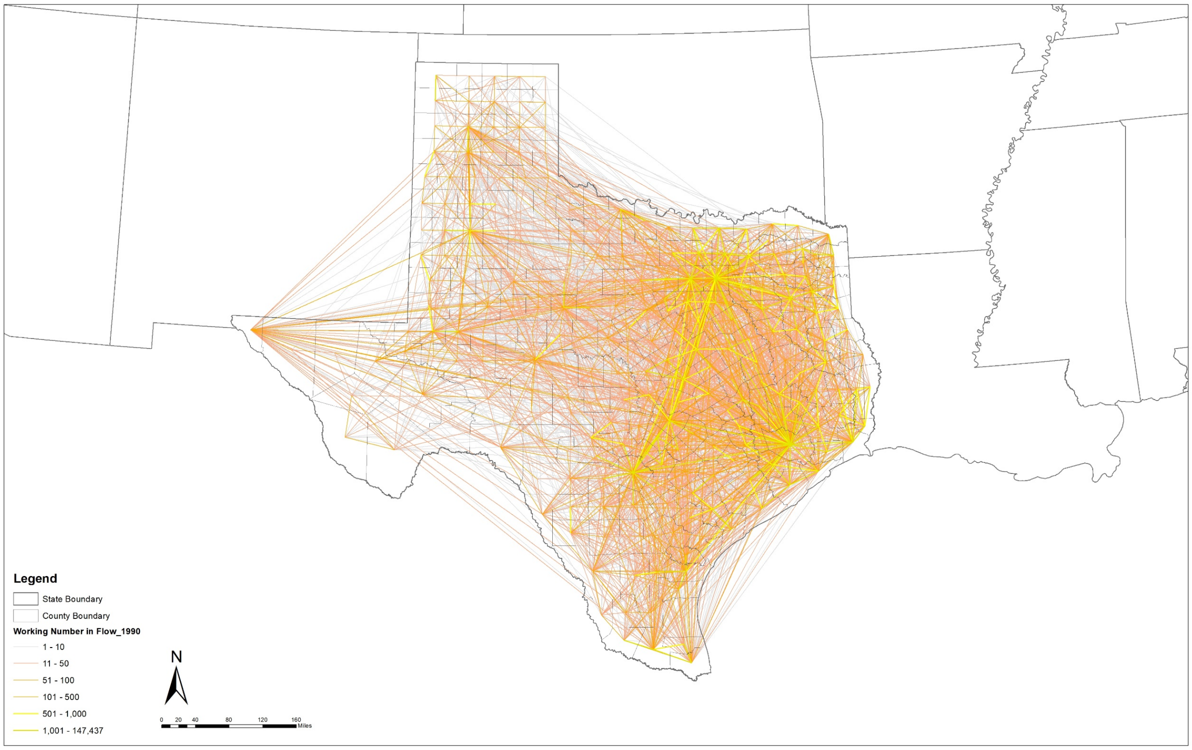

Figure.3. County-to-County Commuting Flows in 1990, Texas

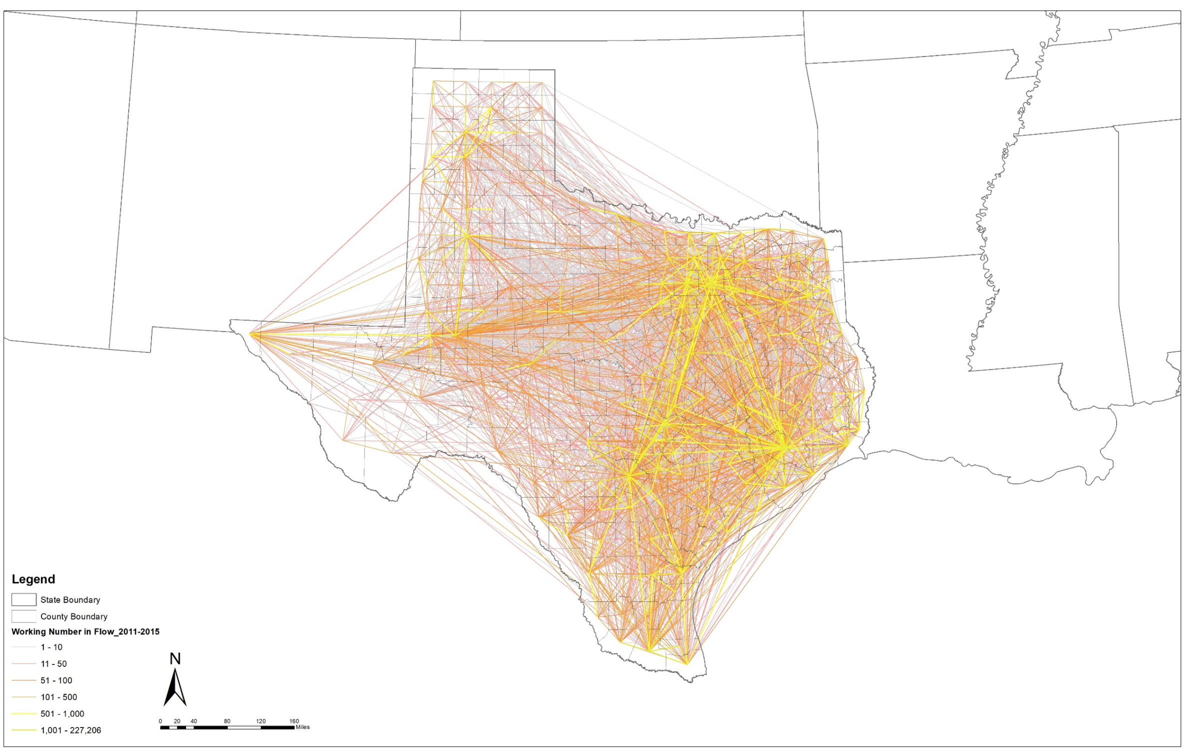

Figure.4. County-to-County Commuting Flows during 2011-2015, Texas

4.2 General Statistical Overview for Inter-County Commuting

In 1990, the total number of commuters with their residence counties and workplace counties in Texas was 7.6 million. During 2011-2015, the number reached 11.9 million. In 1990, the vast majority (83.8%) of the commuters lived in the same county as where they worked. The remainder of the commuters (16.2%) commuted to work in a different county. Compared to 1990, the share of commuters living and working in the same county has declined to 78.3% during 2011-2015. Commuters are becoming less restricted in the choices of their workplaces and exposed to more job opportunities outside the residence county.

Figure.5 and *Figure.6* presents the top 10 counties as workplaces with the largest number of commuters in 1990 and during 2011-2015 respectively. Over the past two decades, Harris County, Dallas County, Tarrant County, Bexar County and Travis County kept their rankings the same. Harris County leads among the counties by 1.5 million commuters in 1990 and 2.5 million commuters during 2011-2015. Dallas County follows with 1.2 million commuters in 1990 and 1.7 million commuters during 2011-2015. As shown in *Figure.7* and *Figure.8*, most of counties, as workplaces with large number of commuters, belong to metropolitan areas within Texas Triangle, such as Houston, Dallas-Fort Worth, San Antonio, and Austin metros. This pattern of distribution becomes more obvious over the targeted time span.

Table.2 shows ten fastest-growing counties as workplaces in Texas over the past two decades, which indicates that the state's suburban counties within large metros are home to the most rapid growth of commuters. The percentage change of Collin County, Williamson County, Loving County, Montgomery County and Rockwall County have exceeded 200 percent and these

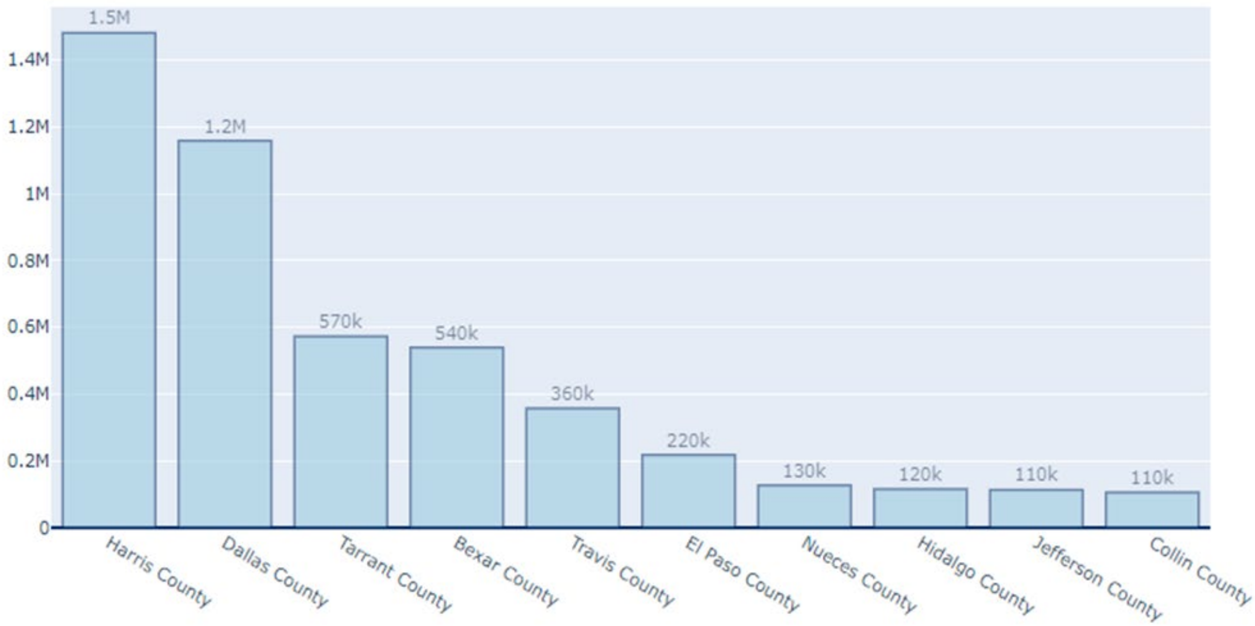

Figure.5. Top10 Counties as Workplaces with the Largest Number of Commuters in 1990

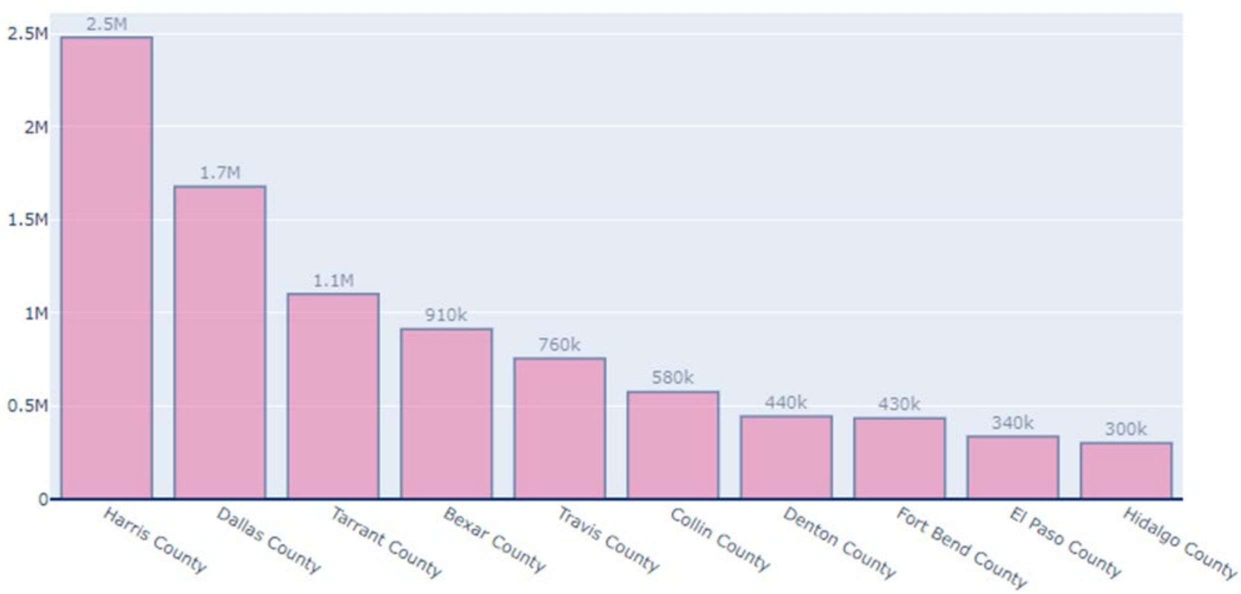

Figure.6. Top10 Counties as Workplaces with the Largest Number of Commuters during 2011-2015

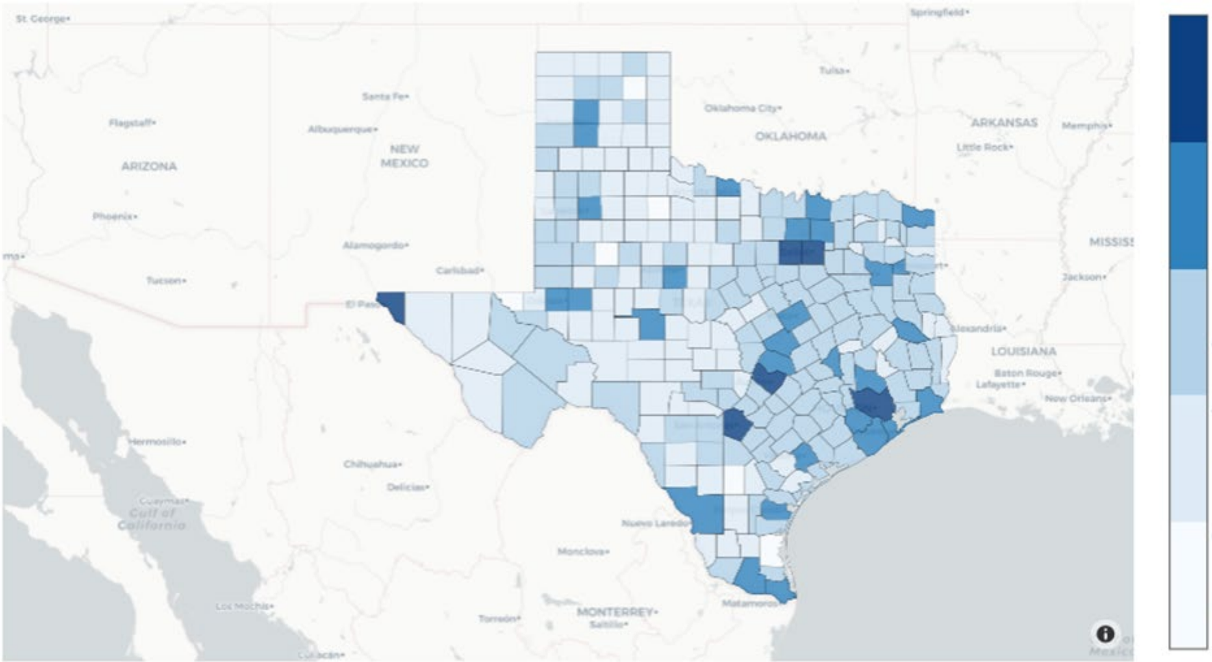

Figure.7. Number of Commuters (Log) in County as Workplace in 1990

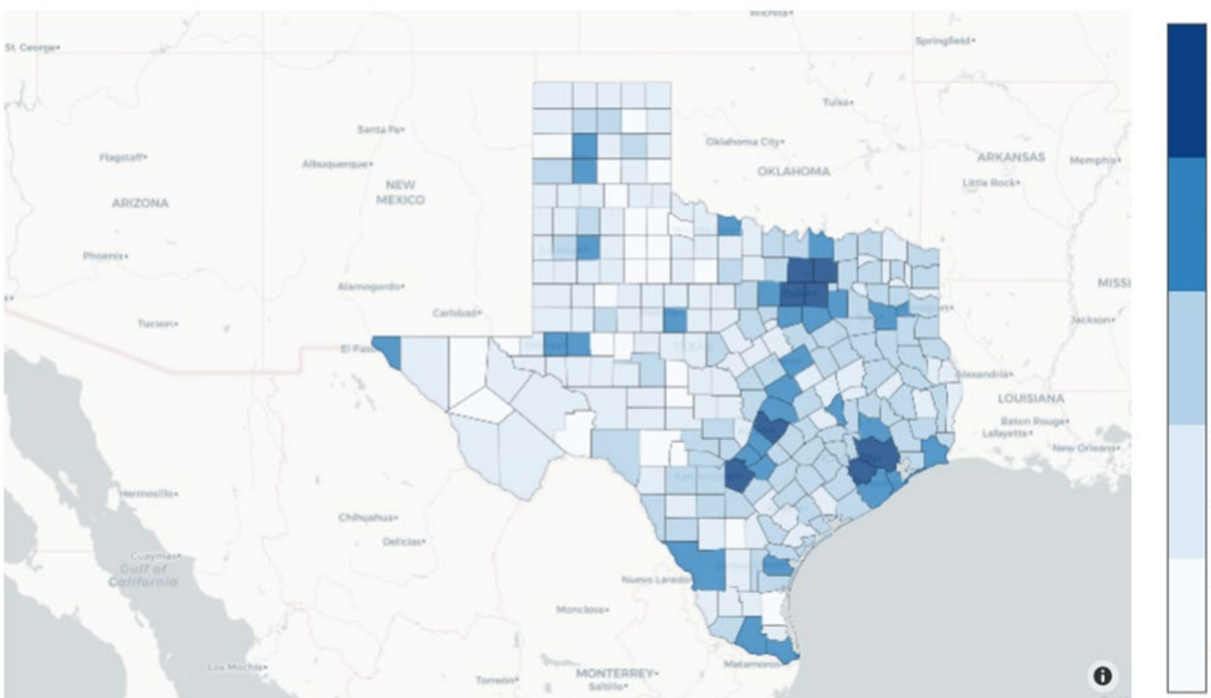

Figure.8. Number of Commuters (Log) in County as Workplace during 2011-2015

County	1990 Commuters	2011-15 Commuters	Percentage Change
Collin County	105,984	386,865	**265.02%**
Williamson County	47,958	167,350	**248.95%**
Loving County	69	238	**244.93%**
Montgomery County	51,682	177,377	**243.21%**
Rockwall County	9,918	32,966	**232.39%**
Denton County	81,766	244,324	**198.81%**
Hays County	22,216	65,756	**195.98%**
Kendall County	4,898	13,947	**184.75%**
La Salle County	1,770	4,961	**180.28%**
Dimmit County	3,098	8,551	**176.02%**

Table.2. Percentage Change of Commuters in Counties as Workplaces

counties are mainly surrounding the state's five largest cities – Houston, San Antonio, Dallas – Fort Worth and Austin. When considering each county in Texas as the workplace of commuters, the commuters in counties of four largest metro areas - Houston, Dallas-Fort Worth, San Antonio, and Austin - account for 54.5 percent, and 58.7 percent of all commuters, in 1990 and during 2011-2015 respectively. This noticeable increase also demonstrates that people are more likely to work in large metros rather than small areas.

Meanwhile, suburban counties surrounding the state's five largest cities - Houston, Dallas, Fort Worth, San Antonio, and Austin - have significant chunks of their population working outside the home counties. In 1990, commuters who lived in most of the suburban counties and worked in central counties account for more than one-third, and even more than one-half, of suburban counties' workforce. *Figure.9* presents the percentage of suburban counties' workforce that commute into their central counties. (In this study, the workforce refers to workers 16 years and older commuting to work in Texas before the annual survey was conducted in 2015):

1. Harris County (Houston): Among suburban counties, Fort Bend County sent the highest share of commuters into a nearby urban county in 1990 with 62.5 percent of its commuters - or 67,372 commuters - heading into Harris County, home to Houston. Commuters commuting from Fort Bend County to Harris County even incredibly outnumbered the intra-commuters of Fort Bend County, which was 37,106.

2. Dallas and Tarrant Counties (Dallas/Fort Worth): Neighboring Dallas and Tarrant counties, home to Dallas and Fort Worth, shared commuters from surrounding counties. In 1990, Tarrant County sent the largest number of commuters to Dallas, which was 104,418.

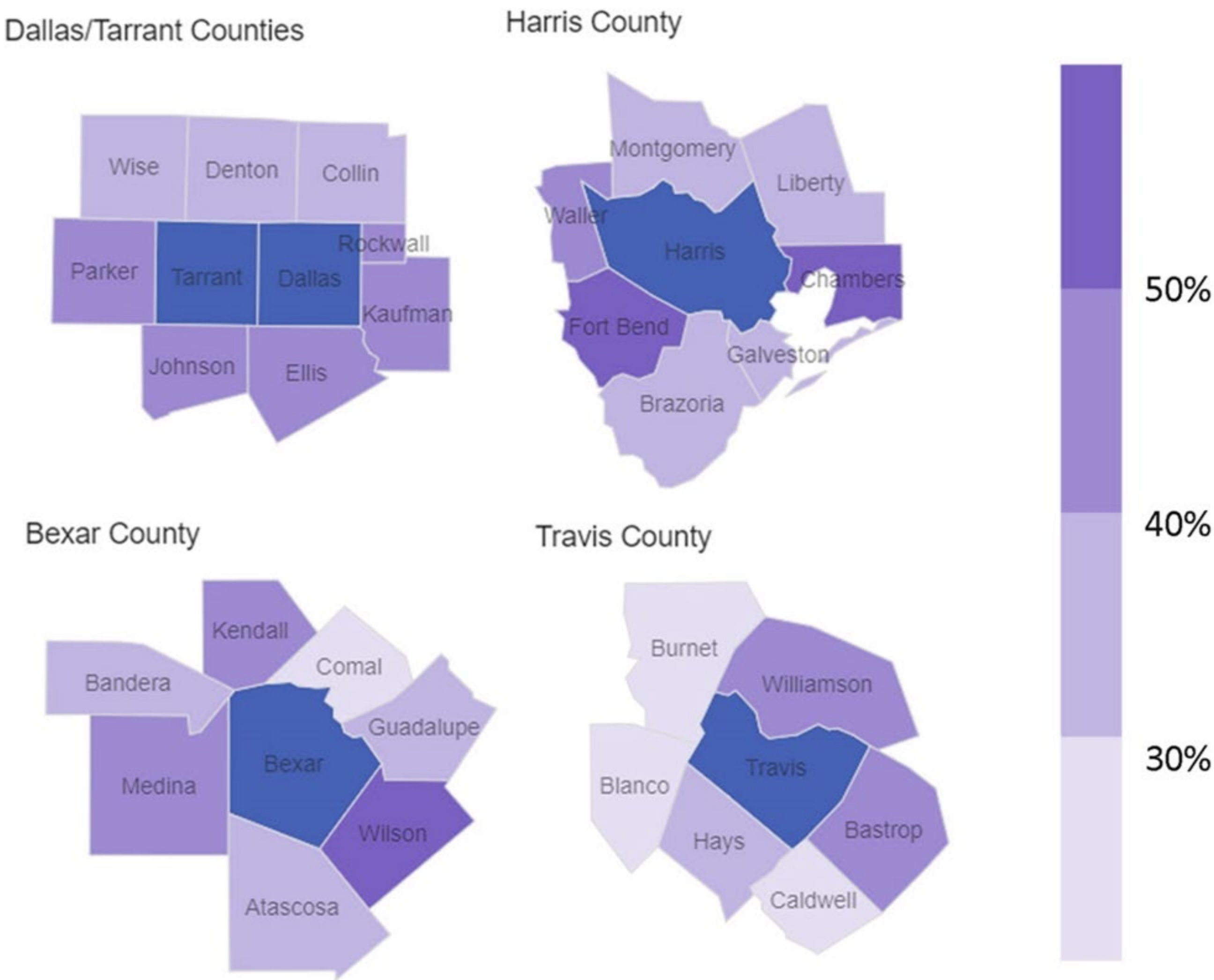

Figure.9. Percentage of Workforce that Commutes to Highlighted Counties in 1990

3. Bexar County (San Antonio): The highest share of commuters into Bexar County, home to San Antonio, came from Guadalupe County in 1990, which sent around 51 percent of its

workforce - or 4,722 commuters. Guadalupe County sent the highest raw number of workers – 8,765 commuters — into Bexar County.

4. Travis County (Austin): Williamson County accounts for the biggest share of commuters sent into Travis County where Austin is located, which was nearly 56.4 percent of its commuters into the urban county in 1990. Meanwhile, Williamson County had the highest raw number of commuters with 39,687 of its residents working in Travis.

During 2011-2015, more than a million Texans living in suburban counties near the five largest Texas cities commute into the metropolitan areas. Similar to 1990, in most of the suburban counties, these commuters account for more than one-third, and even more than one-half, of their workforce. *Figure.10* presents the percentage of suburban counties' workforce that commute into their central counties. (In this study, the workforce refers to workers 16 years and older commuting to work in Texas before the annual survey was conducted in 2015):

1. Harris County (Houston): Among suburban counties, Fort Bend County sent the highest share of commuters into a nearby urban county during 2011-2015 with 59.1 percent of its commuters - or 181,752 commuters - heading into Harris County, home to Houston.

2. Dallas and Tarrant Counties (Dallas/Fort Worth): Neighboring Dallas and Tarrant counties, home to Dallas and Fort Worth, shared commuters from surrounding counties. During 2011-2015, Collin County surpassed Tarrant County, sending the largest number of commuters to Dallas, which is 152,920. Tarrant County had 144,079 commuters heading to Dallas County during 2011-2015.

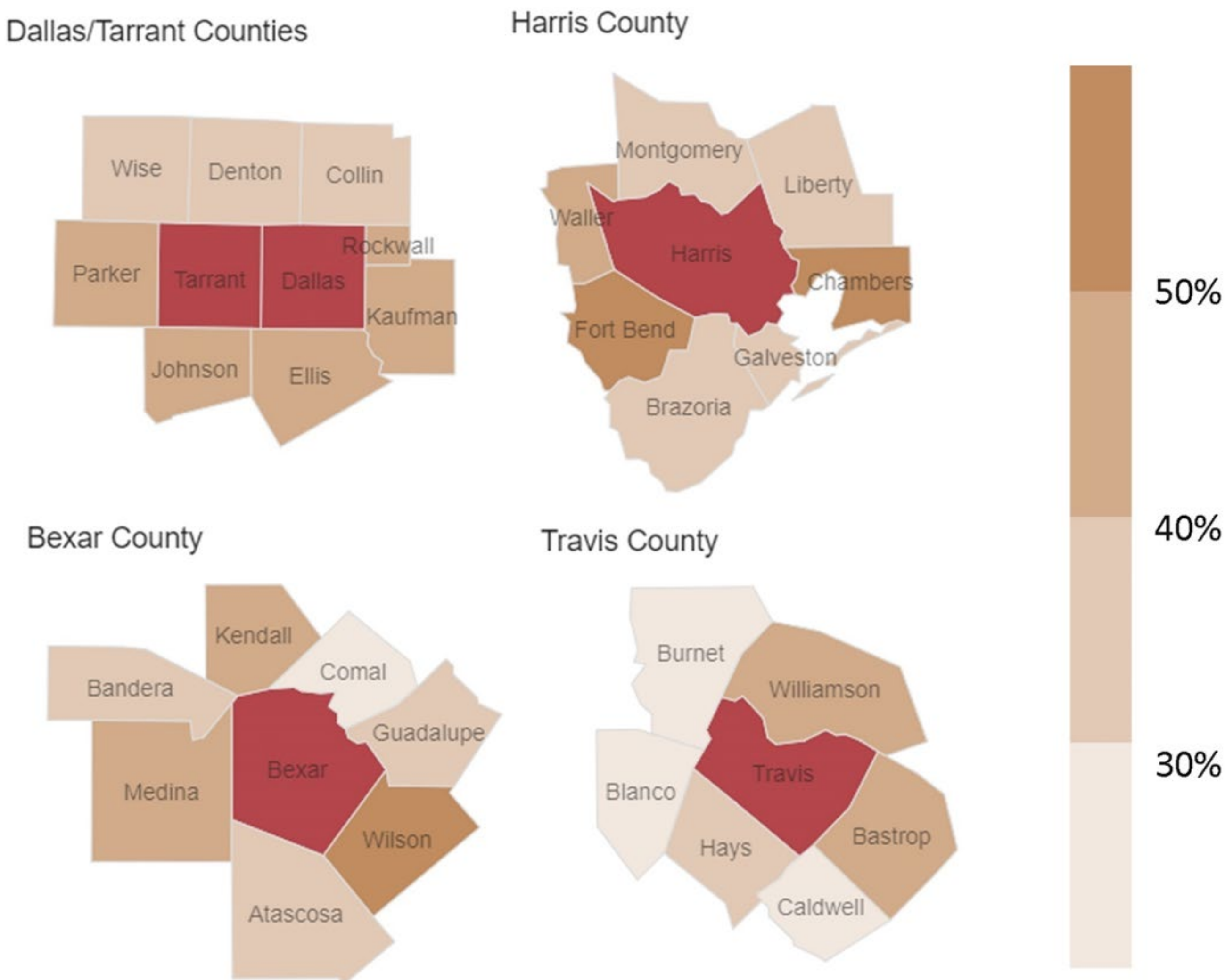

Figure.10. Percentage of Workforce that Commutes to Highlighted Counties during 2011-2015

3. Bexar County (San Antonio): The highest share of commuters into Bexar County, home to San Antonio, came from Wilson County during 2011-2015, which sent around 55 percent of its

workforce - or 9,781 commuters. Guadalupe County sent the highest raw number of workers – 23,169 commuters — into Bexar County.

4. Travis County (Austin): Bastrop and Williamson counties nearly tied for the biggest share of commuters sent into Travis County where Austin is located, which were both nearly 45 percent of their commuters into the urban county during 2011-2015. But Williamson County had the highest raw number of commuters with 99,787 of its residents working in Travis.

Tables 3,4,5,6 summarize the changing trends of how workforce of suburban counties commuted to core counties in metropolitan areas. Overall, in Texas, the percentages of out-of-county workers are not significant, most of them are located outside the triangle region in 1990. During 2011-2015, the percentages have grown in the triangle region. It is not a surprise that some suburban counties witnessed more than half of their workforce heading for jobs in the core county. With growing number of workers looking for affordable housing in the suburbs, long-distance commute or mega-commute will keep largely increasing in the near future.

4.3 The Trend of "Super Commute"

Currently, metropolitan regions are rapidly growing. Because of being building around highway system, metropolitan areas are facing quality of life constraints, such as traffic congestion, loss of open space, and rising greenhouse gas emissions (Todorovich, 2009). Both the shape and behavior of metropolitan areas are changing in a fast pace as well. The interregional commuting is rising because people and jobs are migrating outwards. The sprawling edge cities form adjacent regions that are sharing strong economic ties, overlapping commuting patterns and frequent

Dallas County (Dallas)			
Year / Suburban Counties	1990	2011-2015	Percentage Change
Rockwall County	7,380	18,161	146.08%
Kaufman County	9,655	21,602	123.74%
Collin County	71,044	152,920	115.25%
Wise County	645	1,224	89.77%
Denton County	66,720	110,780	66.04%
Ellis County	15,574	24,757	58.96%
Parker County	1,204	1,882	56.31%
Tarrant County	104,418	144,079	37.98%
Johnson County	3,044	4,002	31.47%

Table.3. Changes of Workforce in Suburban Counties Commuting to Core County (Dallas County)

Harris County (Houston)			
Year / Suburban Counties	1990	2011-2015	Percentage Change
Brazoria County	21,167	58,805	177.81%
Fort Bend County	67,372	181,752	169.77%
Montgomery County	36,769	87,418	137.75%
Waller County	3,472	8,051	131.88%
Chambers County	3,369	7,753	130.13%
Galveston County	26,984	50,755	88.09%
Liberty County	6,506	9,734	49.62%

Table.4. Changes of Workforce in Suburban Counties Commuting to Core County (Harris County)

Bexar County (San Antonio)			
Year / Suburban Counties	1990	2011-2015	Percentage Change
Guadalupe County	8,765	23,169	164.34%
Kendall County	2,474	6,176	149.64%
Medina County	3,593	8,487	136.21%
Comal County	6,012	13,941	131.89%
Wilson County	4,722	9,781	107.14%
Atascosa County	3,739	7,131	90.72%
Bandera County	1,631	2,977	82.53%

Table.5. Changes of Workforce in Suburban Counties Commuting to Core County (Bexar County)

Travis County (Austin)			
Year / Suburban Counties	1990	2011-2015	Percentage Change
Hays County	10,504	31,690	201.69%
Williamson County	39,687	99,787	151.43%
Burnet County	847	1,803	112.87%
Bastrop County	7,948	13,132	65.22%
Caldwell County	2,955	4,625	56.51%
Blanco County	353	493	39.66%

Table.6. Changes of Workforce in Suburban Counties Commuting to Core County (Travis County)

business travel.

"Super Commuter" is becoming a new norm and its trend has been consistently increasing over time. Moss and Qing's study (2012) defined a "super commuter" as a person who works in the central county of a given metropolitan area, but lives beyond the boundaries of that metropolitan area. With the changing employment landscape, super commuters are commuting long times and distances to get to work by air, rail, car and any combination of modes. According the U.S. Census Bureau, metropolitan area boundaries are based on the degree of "social and economic integration, as measured by commuting to work" between adjacent areas and the relative urban core. It is not doubt that the number of super-commutes is on the rise since the labor sheds (where workers live) has been expanding and commuting patterns are becoming more and more interregional (Moss and Qing, 2012). One of the results from interregional commuting patterns is the definition of metropolitan geographies needs to be reconsidered and even new higher level of geographic unit should be invented. In order to examine the overall trend of super-commute in Texas over time, this study delineates 50-mile (one-way flow) as a threshold between regular commuting activities and super commuting flows, which refers to the definition from the report of U.S. Census Mega Commutes (2013). The calculation for county-to-county centroid distance was operated in ArcGIS and each individual home-to-work flow pair is based on Euclidean distance.

It turns out that Texas has become the "epicenter" for super-commute growth over the past decades. In Texas, the number of super commuters had exceptionally increased from 405,384 in 1990 to 931,030 during 2011-2015, at approximately 130%. Across the state, the city labor sheds are expanding rapidly, and the growth rate of super commuters are far outpacing workforce growth rates. For both 1990 and during 2011-2015, the largest super-commute flow is Tarrant County-Collin County, from the volume of 4,902 to 15,386 commuters. Approximately 14 percent of the workforce of Harris and Dallas Counties lived outside the combined metropolitan areas. Meanwhile, Dallas led the list of super-commute counties in the state, with those commuters from Austin and Houston. Furthermore, counties within in the "Texas Triangle" corridor features the fastest growing pace over the past decades. *Figure.11 and Figure.12* highlights how the number of super commuters among four largest metro areas has increased, along with *Table.7* statistically presenting the change of inter-metro commutes over the past two decades. The number of commuters between metro areas of Austin-Round Rock-Georgetown and San Antonio-New Braunfels nearly tripled. It is interesting to notice that the largest three increases of inter-metro commutes are related to Austin-Round Rock-Georgetown metro area. This is might because of the booming of the city of Austin, which had contributed to the settlement of numerous industries and people. Moreover, all of the super commutes along "Texas Triangle" corridor had increased more than at least 50 percent over the past two decades. The reasons for supper commute might be much complicated: some are escaping from unaffordable housing, resulting in job-housing separation; some are willing to move with the purpose of better education for kids; while others are looking for real estate interests out of certain areas. Zillow (2015) mentioned that in tech job centers like Seattle and San Francisco, low-income workers are moving farther and farther outwards while high-income workers are able to still afford to live close to their workplaces.

Over the past 20 years, the geographic proximity played a less relevant role in precondition for metropolitan areas because of the technological advances, such as mobile communications and teleconferencing (Moss and Qing, 2012). In Texas case, some emerging "Texas Triangle" cities can be apart more than 200 miles from each other, but the trends towards urban integration and "super-commute" are unstoppable. Evolving in the information age, such social and economic activities in Texas cities are becoming increasingly inter-regional. In order to be competitive

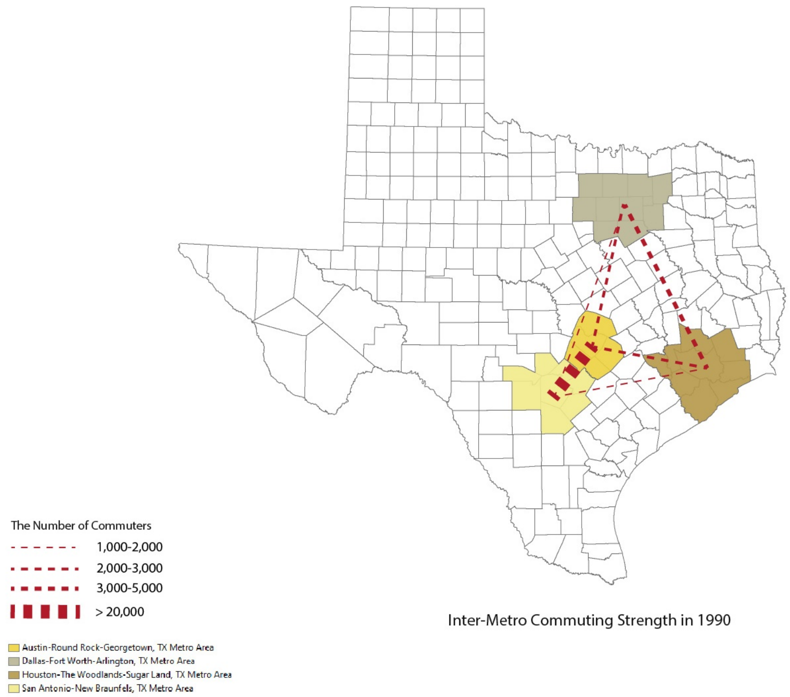

Figure.11. Inter-Metro Commuting Strength in 1990

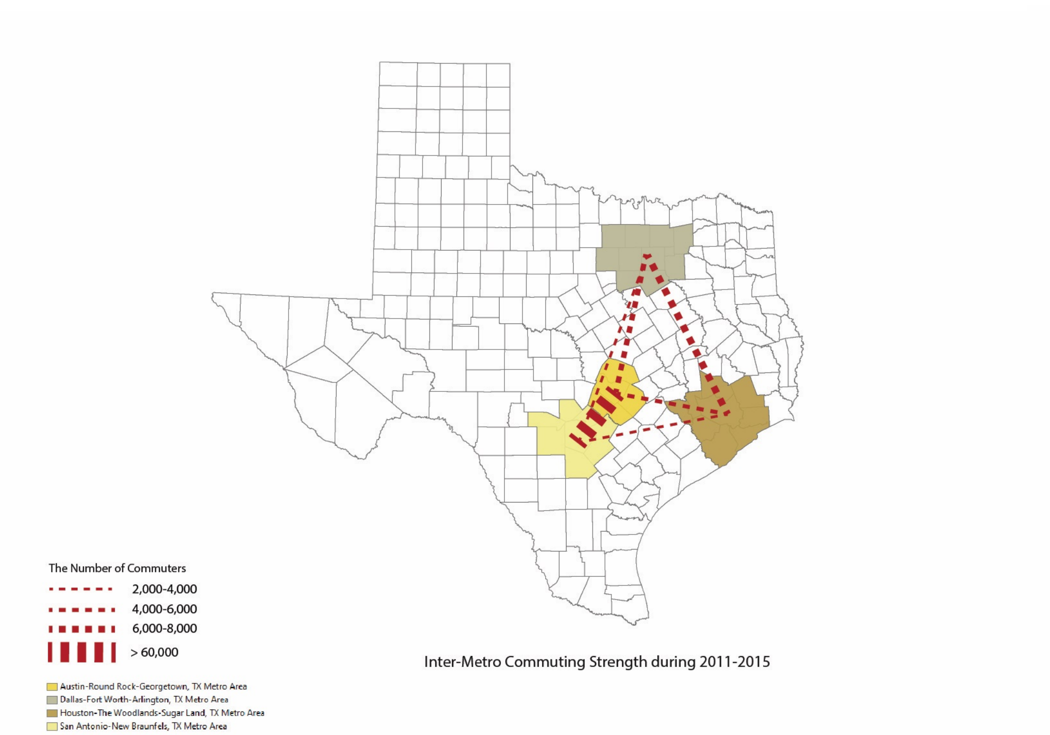

Figure.12. Inter-Metro Commuting Strength during 2011-2015

	1990	2011-2015	Percent Change
Austin-Round Rock-Georgetown and San Antonio-New Braunfels	21,161	63,504	**200.10%**
Dallas-Fort Worth-Arlington and Austin-Round Rock-Georgetown	2,235	5,204	**132.84%**
Austin-Round Rock-Georgetown and Houston-The Woodlands-Sugar Land	3,009	6,735	**123.83%**
Dallas-Fort Worth-Arlington and San Antonio-New Braunfels	1,396	2,806	**101.00%**
Houston-The Woodlands-Sugar Land and San Antonio-New Braunfels	1,928	3,336	**73.03%**
Dallas-Fort Worth-Arlington and Houston-The Woodlands-Sugar Land	4,505	7,175	**59.27%**

Table.7. Commuters Change among Four Largest Metro Areas of Texas

in the whole nation's or even global economy, planning strategies for "mega-regions" are necessary in the upcoming future. As presented above, the integrated mega-region will play an increasingly important role as driving force of economic growth in both advanced and emerging nations (Moss and Qing, 2012).

4.4 Algorithmic Community Partitioning

The establishment of commuting flow network, the GN algorithm, the calculation of edge weights and scores of weighted modularity, were implemented using Python and its open-source packages. In the first run of the 2011-2015 ACS data set limited to commutes both originating and concluding within the state of Texas, the Girvan-Newman algorithm could produce an ideal result by assigning the weight of each edge (flow) as its original volume of commuters. A large number of communities were in line with what are presented in visual flow mapping of 2011-2015. However, the initial output communities also exhibited a considerable amount of meaningless clusters when evaluated visually. Certain counties were assigned into communities which displayed little or no geographic sensibility. After several runs of tests, a more efficient and accurate partitioning was achieved. Instead of simply defining each edge's weight as the original volume of commuters, the linkage coefficient was introduced to be reconsidered as the weight of edge as previously mentioned. That means, the larger the proportion of the total commuters between two counties, the stronger is the bond between two counties.

The computation process of the GN algorithm is relatively slow, which is restricted to the number of nodes in the assigned network. As a result, the GN algorithm produced a hierarchical clustering from top and down by taking" linkage coefficient" as weight. For every time of clustering, the weighted modularity was calculated. The results consist of the weighted modularity score with a different number of communities detected. *Figure.13* shows the weighted modularity reaches its highest score while the number of clusters falls in 35, associated with *Figure.14* presenting the result of county-level clustered data (each node can be counted as county centroid). The weighted modularity score is 0.29. This generated the most successful partitioning in areas where nodes (county centroids) are quite well-linked in the data set to many other nodes, such as

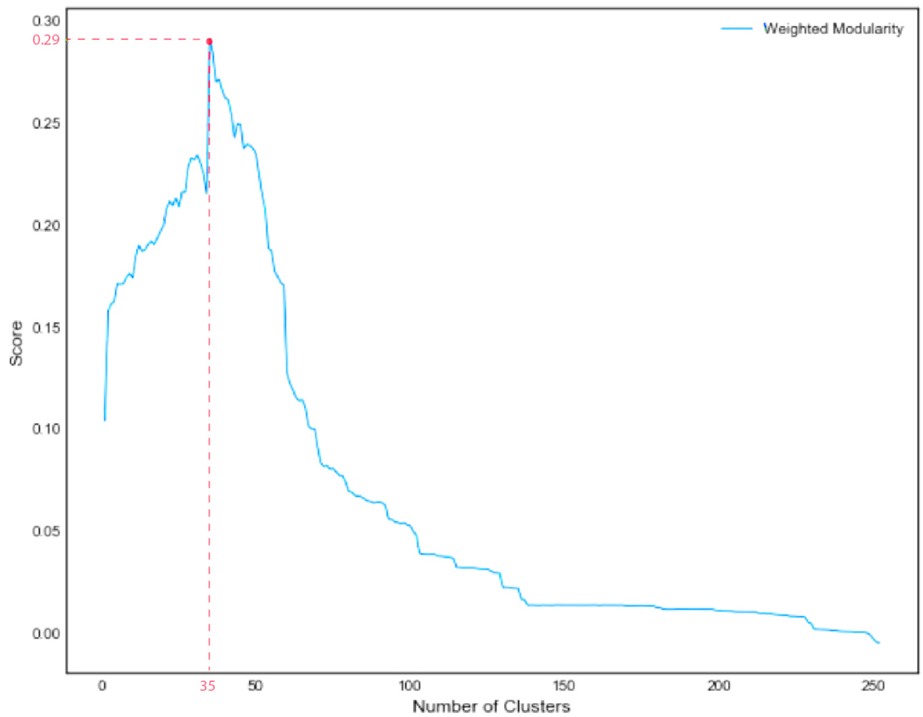

Figure.13. Weighted Modularity Score for Every Number of Clusters

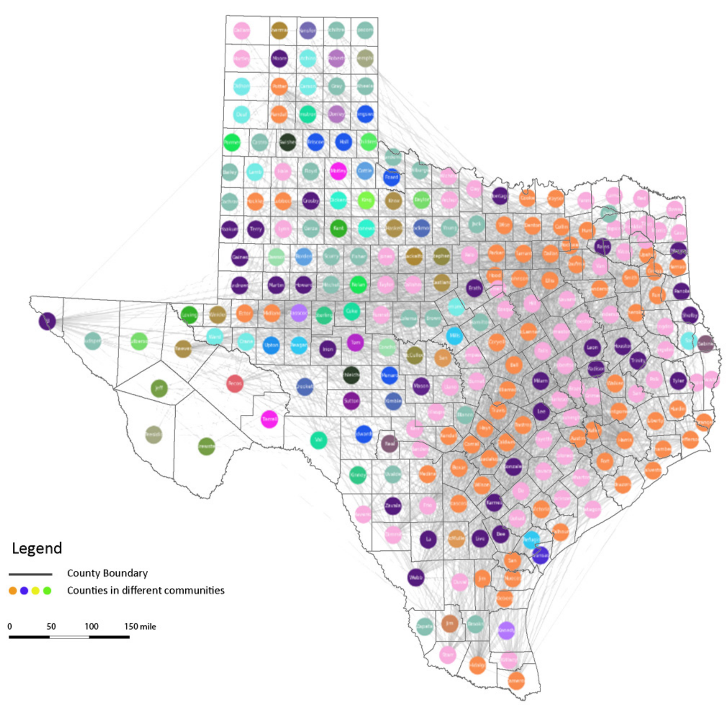

Figure.14. County-level Clustered Data in Texas

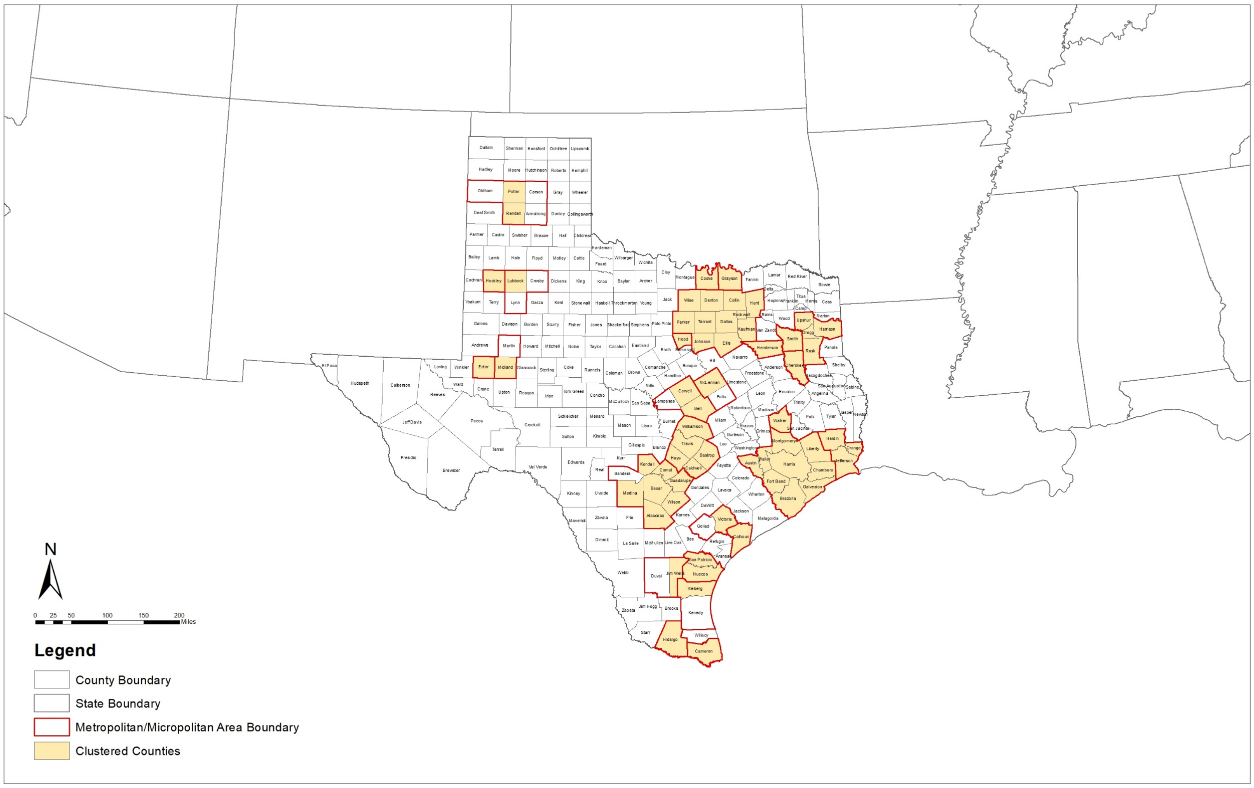

Figure.15. Major County Clusters in Texas

was found in large metropolitan areas. As is evident in *Figure.15*, the GN algorithm employed in the study was able to divide the state of Texas into multiple geographically – continuous or discontinuous regions, which includes one largest “mega-community”. This “community” could be interpretively recognizable as “mega-agglomerations”, with major counties (cities) at each continuous region’s center: for example, Dallas County (City of Dallas), Harris County (City of Houston), Bexar County (City of San Antonio) and Travis County (City of Austin). Based on their geographic proximity, the largest “community” is primarily presented as a triangular geometry. This offers a strong evidence that county-level commuting patterns certainly do divide space into different clusters of regional labor market, and that the general structure of "mega-agglomeration" or “mega-region”, is able to be detected algorithmically. Furthermore, *Figure.15* also depicts how clustered counties form the connectivity between metropolitan and micropolitan areas. The highlighted borders show the places where major MSAs are situated. It is noticeable that the largest county cluster detected by the GN algorithm in the study correspond very closely to the major MSAs defined by census in Texas. Formation of the largest county cluster with correspondent MSAs yields the result in *Table.8.* Metropolitan areas of Dallas-Fort Worth-Arlington, Houston-Woodlands-Sugar, San Antonio-New Braunfels, and Austin-Round Rock-Georgetown, dominate the list, consisting of more than half number of counties within this cluster. The algorithmic result for finding spatial clusters is highly consistent with visual interpretation of the ACS 2011-2015 data set. Other MSAs that have less population and smaller geographic scale, such as Longview, Tyler, Waco metro areas, are also being integrated into this "mega-community".

Additionally, integration has been witnessed within Western and Southern parts of Texas as well. Several counties within MSAs of Western Texas, such as Midland County, Lubbock County, and Randall County, emerge on the list. Other than those, Counties along the Gulf Coast also seem to be integrated algorithmically, like Nueces County and Cameron County. All of these demonstrate that the commuting landscape have stretched from the triangle region to the outside parts of Texas's core areas. Under a broader context, the Gulf Coast region in Texas has one metropolitan area: the Houston-The-Woodlands-Sugar Land MSA. Harris County, with the city of Houston as its center, is the economic hub of the region. As is shown in the results from algorithmic approach, several counties in the Gulf Coast region, besides Houston metro area, are associated with the core clustered areas. Regional interconnection in Texas is once again demonstrated. According to The Gulf Coast Regional Report (2018), the region took up 25 percent of the state's population by the end of 2017, growing 16 percent since 2010, and it is the most diverse region in Texas. Steady economic development has added mor than 400,000 jobs in the region from 2007

MSAs	Clustered County Name	Number of Counties	Percentage of Counties
Dallas-Fort Worth-Arlington, TX Metro Area	Dallas	11	21.57%
	Tarrant		
	Wise		
	Denton		
	Collin		
	Hunt		
	Rockwall		
	Parker		
	Johnson		
	Ellis		
	Kaufman		
Houston-The Woodlands-Sugar Land, TX[1] Metro Area	Harris	9	17.65%
	Fort Bend		
	Brazoria		
	Galveston		
	Chambers		
	Liberty		
	Montgomery		

[1] Table.8. Clustered Counties within Major MSAs. Continued

	Waller[2]		
	Austin		
San Antonio-New Braunfels, TX Metro Area	Bexar	7	13.73%
	Kendall		
	Comal		
	Guadalupe		
	Wilson		
	Atascosa		
	Medina		
Austin-Round Rock-Georgetown, TX Metro Area	Travis	5	9.80%
	Williamson		
	Bastrop		
	Caldwell		
	Hays		
Longview, TX Metro Area	Upshur	4	7.84%
	Gregg		
	Harrison		
	Rusk		
Beaumont-Port Arthur, TX Metro Area	Hardin	3	5.88%
	Orange		
	Jefferson		

[2] Table.8. Clustered Counties within Major MSAs. Continued

Corpus Christi, TX Metro Area	San Patricio	2	3.92%
	Nueces		
Killeen-Temple, TX Metro Area	Coryell	2	3.92%
	Bell		
Brownsville-Harlingen, TX Metro Area	Cameron	1	1.96%
Lubbock, TX Metro Area	Lubbock	1	1.96%
McAllen-Edinburg-Mission, TX Metro Area	Hidalgo	1	1.96%
Midland, TX Metro Area	Midland	1	1.96%
Odessa, TX Metro Area	Ector	1	1.96%
Tyler, TX Metro Area	Smith	1	1.96%
Victoria, TX Metro Area	Victoria	1	1.96%
Waco, TX Metro Area	McLennan	1	1.96%

Table.8. Clustered Counties within Major MSAs

to 2017. Extending from Texas' part, the development of the Gulf Coast also owes much to the adjacent state of Louisiana. The rapid growing interconnection between two states-Texas and Louisiana-along the Gulf Coast will undoubtedly encourage increasing number of cross-state commuting activities in the near future. Hence, not only cross-regional planning strategies, but also even cross-state planning policies are supposed to be fully considered and implemented accordingly.

4.4.1 Discussion for algorithmic approach

Using the ACS 2011-2015 Commuting Flow data set, the study explored the application of megaregional delineation through algorithm-based community detection methods. The results are highly consistent with the visual interpretation and descriptive statistics from the data set. However, there are still several limitations in the algorithmic approach so far. First, the GN algorithm requires relatively small network structure. This means it is not time efficient with networks containing large number of nodes (over thousands of nodes). Therefore, the GN algorithm might be only applicable in urban and regional studies when it comes to the same spatial context as in this study. Second, the factor of spatial distance is not taken into consideration. Counties that are close to each other should be more integrated in terms of social interaction, economy, and culture. Without considering spatial distance, the highly valued "spatial proximity" in regional science is totally ignored. All nodes are counted as indistinguishable points, which are in contrast to that in reality. Last but not the least, the parameter setting serves as the most important role leading the algorithmic result. The accuracy of the application of the community detection algorithm is rooted in its parameter setting (Wu et al., 2019). In order to achieve high performance of the community detection algorithm, researchers are not supposed to rest on subjective knowledge and empirical judgment.

There is no doubt that four largest metropolitan areas in Texas drive the state's economy: Houston, Dallas – Fort Worth, San Antonio and Austin and these largest metro areas are credited for 85 percent of the state's overall population growth since 2010 (Fulton, 2019). They have most of the state's jobs, and are driving population growth in dramatic fashion. However, by algorithmic approach developed in this study, it is not a surprise to observe that the largest cluster in the most successful community partitioning is highly consistent with the geographic distribution of major metro areas in Texas by comparing *Figures.15 and 16*. Furthermore, the largest cluster do exist to indicate that these major metro areas are not separate. In other words, they function as one integrated entity in Texas because of their interconnected economic activities with each other, which can be named as a "megaregion". This megaregion develops as a geometric form of triangle in the central and eastern part of the state., anchored by four largest metropolitan areas, and associated with other small dispersed metro areas. It is worth believing that this triangular megaregion will continue its role as Texas economic engine, or even the economic engine of the southwestern United States. Although the cores of this megaregion are miles apart and its development is either not linear or contiguous, they remain physically close enough so that mutual competition and cooperation have forced them and continue enabling them to seek out different and complementary economic roles. This engine holds four major cylinders to power the Texas economy.

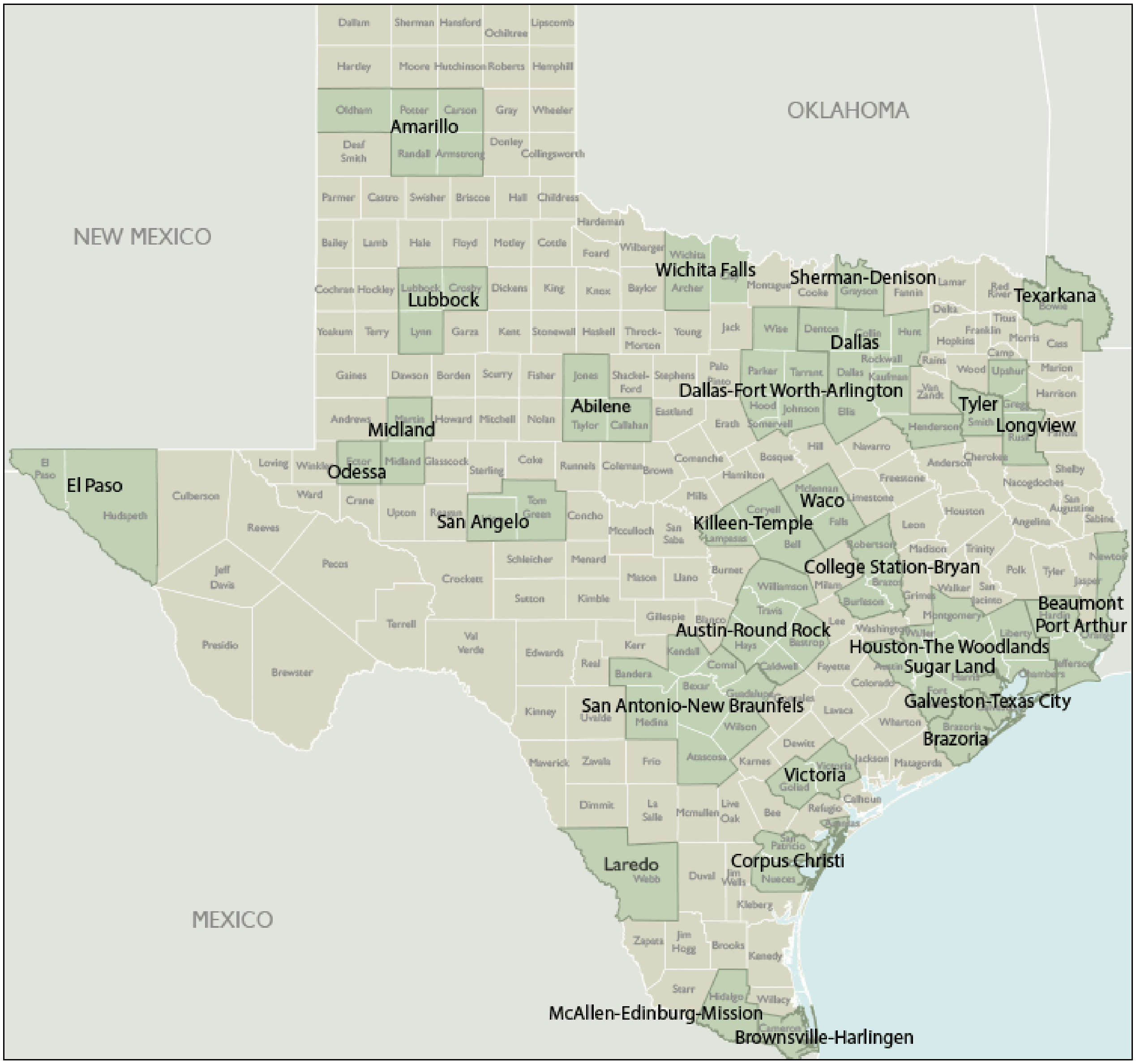

Figure.16. Metropolitan / Micropolitan Areas in Texas

Chapter 5: Summary and Conclusions

As all of the suggested in discussion of the previous chapter, large metropolitan areas in Texas play their complementary roles and, in the meantime, come together to form a great economic entity or engine that serves the Texas. A clear-cut, coherent spatial structure is exhibited by the geography of labor and commuter interaction in Texas, which is outlined by Dallas-Fort Worth (DFW), Houston, San Antonio, with Austin locating at the Dallas-San Antonio line. It is also demonstrated by geography-blind algorithmic analysis of community structure in Texas commute data set, that the four large metropolitan areas - Houston, Dallas-Fort Worth, San Antonio, and Austin - associated with other relatively small metro areas, in terms of population, geographic scale, and etc., are classified as one "community". This means, economic interaction happening among these areas enables them to reach a high level of agglomeration or integration, distinguishing them from other parts of Texas. Meanwhile, this mega "community" exist as a dominant form of triangle resulting from spatial proximity of clustered counties or metro areas in the "community". These conclusions are evident in the first visual interpretation and statistical method as well. It turns out that both methods developed in the study provide an expected similarity in terms of the economic geography of Texas, and support the assumed existence of a megaregion in Texas.

A megaregion does exist as a coherent, powerful unit of economic activity. In this case, it is the Texas Triangle. Such an entity requires new large-scale strategies or policies in various fields, including transportation planning, environmental planning, real estate, infrastructure development and etc. Megaregional planning can enhance current metropolitan and city level planning for economic development, infrastructure investments, environmental protection, and rural and urban land uses (Zhang et al., 2007). Based on the analysis of this study, the author suggests that increased transportation investment and expansion into new modes of transportation (especially rail) would be necessary with the purpose of ensuring the continued economic competitiveness of megaregion, in this case, Texas Triangle. Existing clusters of major metropolitan areas in Texas can make transportation planning on megaregional scale logical and appropriate. The interface between transportation investment and economic development has extended implications which undoubtedly go beyond transportation's basic purpose of moving goods and people from one place to another (Eberts, 2016). An efficient transportation system can improve the productivity of the economy. In the United States, understanding the effects of attributes of highways on economic development has been largely explored. But do we know much about the effects on productivity of reducing highways congestion or expanding new modes of transportation like high-speed rail? All of these innovations can be developed under the megaregional context. However, policy makers and practitioners need to gain a clear picture of the effects of these innovations on economic development through enhanced development of transportation services and a more efficient use of scarce resources. Scarce resources may extend beyond transportation investment to include land use, air quality, and etc. (Eberts, 2016).

Overall, the study initiates a first step towards a better understanding of regional economic geography of Texas. Organizations, such as Regional Plan Association, the Federal Highways Administration might value the similar methods to take deeper nationwide delineation of megaregions and furtherly prepare for the future regional planning.

Appendixes

Python scripts for this study are attached as below. Contents included: the network establishment for the ACS 2011-2015 County-to-County Commuting Flows; calculation of the GN algorithm; and calculation of the weighted modularity scores.

Creating Inter-County Commuting Network

In [45]:
```
import pandas as pd
import networkx as nx
import matplotlib.pyplot as plt
import numpy as np

path1 = r'D:\UT Austin\GRA-Dr. Zhang\intercounty_commute\2015TexasCommuting.xlsx'
path2 = r'D:\UT Austin\GRA-Dr. Zhang\intercounty_commute\TexasCentroidstxt.xlsx'

df_edge = pd.read_excel(path1,'Sheet1')
df_node = pd.read_excel(path2)


df_edge_1 = df_edge.sort_values('WorkingNum',ascending=False)

T = nx.Graph()

edges_weight = []
for i,n_row in df_edge_1.iterrows():
    edges_weight.append((int(n_row[1]),int(n_row[2]),n_row[-1]))
T.add_weighted_edges_from(edges_weight)
T.add_node(48007)

for i, n_row in df_node.iterrows():
    if n_row['Id'] in list(T.nodes()):
        T.nodes[n_row['Id']].update(n_row[1:].to_dict()) # Version of NetworkX

node_positions={}
for i in T.nodes(data=True):
    cor = (i[1]['x'],i[1]['y'])
    node_positions[i[0]] = cor

Labels = {}
for a,b in T.nodes(data=True):
    Labels[a] = b['Label']

pos = nx.spring_layout(T)
plt.figure(figsize=(50,40))
nx.draw_networkx_edges(T,pos=pos,edge_color='black',alpha=0.2)
nx.draw_networkx_nodes(T,pos=pos,node_size=2000,node_color='orange',alpha=0.6)
nx.draw_networkx_labels(T, pos=pos, labels=Labels,font_size='12',font_color='black')
plt.show()
```

Implementing Girvan-Neman Algorithm and Calculating Weighted Modularity Scores

In [5]:
```
from networkx.algorithms.community import girvan_newman
from networkx import edge_betweenness_centrality as betweenness
import random


def most_central_edge(G):
    centrality = betweenness(G, weight='weight')
    return max(centrality, key=centrality.get)

comp = list(girvan_newman(T,most_valuable_edge=most_central_edge))


communities = comp

com = []
for communities in comp:
    com.append(tuple(sorted(c) for c in communities))
```

```python
In [ ]: from igraph import *
import igraph

G = igraph.Graph.DictList(
          vertices=df_node.to_dict('records'),
          edges=df_edge.to_dict('records'),
          directed=False,
          vertex_name_attr='Id',
          edge_foreign_keys=('Source', 'Target'));

resp = {}
for i in G.vs():
    resp[i['Id']] = i.index

MODs = [];NUM = [];count = 1

for i in com:

    community = list(i)

    communitynew=[]
    for x in community:
        cluster = []
        for y in x:
            cluster.append(resp[y])
        communitynew.append(cluster)

    membership = [None]*G.vcount()
    for c, cluster in enumerate(communitynew):
        for v in cluster:
            membership[v] = c

    MOD = G.modularity(membership,weights=G.es['LinkageCoe'])
    MODs.append(MOD)
    NUM.append(count)
    count += 1


plt.style.use('seaborn-white')
fig = plt.figure(figsize=(10,8))

plt.plot(NUM,MODs,'-',linewidth=0.8,label='Weighted Modularity')
plt.xlabel('Number of Clusters',fontsize=12)
plt.ylabel('Score',fontsize=12)
plt.legend(loc='best');
plt.show()

color = []
n = 300
for i in range(n):
    color.append('#%06X' % random.randint(0, 0xFFFFFF))

count = 0

for i in com[35:36]:
    newman = {}
    color_value = []
    for a,b in enumerate(i):
        for i in b:
            newman[i] = a
    for a,b in newman.items():
        color_value.append(color[b])

    plt.figure(figsize=(25,24))
    nx.draw_networkx_nodes(T,pos=node_positions,node_size=1000,node_color=color_value,alpha=1)
    nx.draw_networkx_labels(T, pos=node_positions, labels=Labels,font_size='9',font_color='white')
    nx.draw_networkx_edges(T,pos=node_positions,edge_color='grey',alpha=0.2)
    count += 1
    plt.title('com = {}'.format(count))
    plt.show()
```

References

Acharya, D. (2012). Commuting patterns and labor markets: a new regional classification for Louisiana. LSU Master's Theses. 3279. Available at: https://digitalcommons.lsu.edu/gradschool_theses/3279

Andersson, M., Lavesson, N., & Niedomysl, T. (2018). Rural to urban long-distance commuting in Sweden: Trends, characteristics and pathways. Journal of Rural Studies 59 (2018) 67-77.

Glocker, D. (2018). The Rise of Megaregions: Delineating a new scale of economic geography. OECD Regional Development. Available at: https://dx.doi.org/10.1787/f4734bdd-en

McKenzie, B. (2013). County-to-County Commuting Flows: 2006-10. Census.gov. Available at: https://www.census.gov/library/working-papers/2013/acs/2013- McKenzie.html.

Eberts, Randall. (2016). Understanding the Impact of Transportation on Economic Development. W.E. Upjohn Institute. Retrieved from http://onlinepubs.trb.org/onlinepubs/millennium/00138.pdf

Girvan, M. & Newman, M.E.J. (2004). Finding and Evaluating Community Structure in Networks. Physical Review E 69, 026113.

Hagler, Yoav. (2009). Regional Planning Association: Defining U.S. Megaregions. America 2050.

Lang, E.R & Nelson, A. (2014). Beyond the metroplex: examining commuter patterns at the "megapolitan scale". Lincoln Institute of Land Policy. WP07RL1.

Mccullough, J & Ura, A. (2015). Long Way Home: Census Details Texas Commutes. Retrieved from https://www.texastribune.org/2015/08/27/where-suburban-texans-commute-work/

Ming, Z., Steiner, F. & Bulter, K. (2007). Connecting the Texas Triangle: Economic Integration and Transportation Coordination. The Healdsburg Research Seminar on Megaregions.

Nambiar, Vipin. (2006). Identifying the Texas Triangle Mega Region. The University of Texas at Austin.

Nelson, G.D. & Rae, A. (2016). An economic geography of the United States: From commutes to megaregions. PLoS ONE 2016, 11, e0166083.

Rapino, M & Fields, A. (2013). Mega Commuting in the U.S. Census.gov. Association for Public Policy Analysis and Management Fall 2013 Conference.

Ross, L.C. & Woo, M. (2011). Megaregion and Mobility. Retrieved from: The Bridge, Spring 2011. Urban Sustainability, pp.27-34.

Sandow, E., & Westin, K. (2010). The persevering commuter – Duration of long-distance commuting. Transportation Research Part A 44 (2010) 433 – 445.

Todorovich, P. (2009). America's emerging megaregions and implications for a national growth strategy. International Journal of Public Sector Management. Vol. 22, No.3 (2009), pp.221-234.

Yue, H., Guan, QF., Pan, YT., Chen, LR., Lv, JJ & Yao, Y. (2019). Detecting clusters over intercity transportation networks using K-shortest paths and hierarchical clustering: a case study of mainland China. International Journal of Geographical Information Science.

Wu, K., Tang, JX., & Long, Y. (2019). Delineating the Regional Economic Geography of China by the Approach of Community Detection. Sustainability 2019, 11, 6053; doi:10.3390/su11216053